IJCSN International Journal of Computer Science and Network, Volume 6, Issue 2, April 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

A Personalized Approach to Search Healthcare Videos

[1] Tanvir Ambekar; [2] Dr.Vijaya Musande

[1] Computer Science and Engineering Department , Dr. Babasaheb Ambedkar Marathawada University,
Jawaharlal Nehru Engineering College
Aurangabad , Maharashtra, India

[2] Computer Science and Engineering Department , Dr. Babasaheb Ambedkar Marathawada University,
Jawaharlal Nehru Engineering College
Aurangabad , Maharashtra, India

Abstract - Nowadays, a lot of data available on the web which can be accessed by users to get the required information. Most of the time the user is interested in the more specific information such as healthcare related videos, but such information is not available on the web. In healthcare, the doctors are more enthusiastic to search the videos related to any disease symptoms and their treatment. In a few cases, if the doctors are not able to operate the patient or unable to recognize the root cause of the disease, they examine the previous treatment given to that patient or similar disease patient. Doctors also refer the internet to search the specific video related to that disease by other doctors in order to decide better disease treatment. So, to overcome such issues, making healthcare videos accessible through the video search engine under specific categories for quick retrieval or related search is essential. The proposed system intends to find the relevant videos for a user's query by using the video search engine for healthcare videos. Most of the healthcare videos are easily available or can be recorded at low cost. The proposed system is used to show the most relevant videos by using keyword based label matching for any particular query of a user. In the proposed system, firstly, healthcare videos are collected and speech to text conversion is done to create transcription snippets. Finally, keyword based labeling is done with the help of transcription snippet and the prescription data which is uploaded along with the video. These keywords are also used to categorize the video search results into different groups. The proposed method helps the doctors easily find the specific or relevant video for effective disease prescription analysis in quick time.

Keywords – *Healthcare video retrieval, Video indexing, Speech to text, Prescription tags, Transcription report, Relevance feedback.*

1. Introduction

Information on the web consists of almost everything images, videos, audio and text, most important of it is video as it has audio-visual representation. There are many healthcare videos available on the internet and users can search or view these videos through various existing search engines. As, there are lots of videos present, it is not possible to categorize all the videos and this gets a tiresome job, due to this if a user searches a video or wants just a section or part of the video, he has to look for a whole list of video search results which are elaborated according to the query, sometimes even duplicate videos are present as number of users upload the same video and it becomes really irritating for the user. Example: if a doctor wants to see symptoms of a particular disease or major reason for heart attack, he gets all the videos containing the keyword attack, which is of no use to a doctor as he is finding the video containing information about heart attack. Up to current times there were numerous images, but today there is an increase in the number of videos uploaded on the internet. The search engines are free to use and one can find anything on it.

Keyword based search approach is used by the search engines to search any information on World Wide Web. The search engines have a huge amount of data (audio, video, image or text) and it shows almost all the results for a particular query of user, this becomes useful for searching the documents as nearly repeated data is present in documents but the same is not in the case of videos as a portion of the information which user wants may or may not be present in the result which is shown. Content based video retrieval helps to search the videos effectively. Content-based means the actual content of the video is searched. Content may refer to anything e.g. shape, color, voice, text, etc. There are many video based domains such as YouTube, DAMS, etc. which contain the information or text in html tags. The search engines gather and store data and then ranks them accordingly using some inbuilt parameters. Indirectly inside these search engines indexes

are generated through which the ranking is done and the relevant video is shown. Open directories depend on humans to edit and compile the listings.

The first and foremost step in content based is to segment the video into various image frames, as videos have a motion which is classified into front-end motion and background motion, it becomes easy to segment into image frames, these have many applications such as surveillance, object manipulation, scene composition, and video retrieval. The video is converted into shots. A shot can be defined as an image sequence which has continuous action. These shots are joined at the end to form a video. These can also be considered as the smallest part of the index. Key-frames are nothing but still images that are extracted from the video or shots and have minute information about the shots. They also contain the text of the video. If key frames are extracted correctly or effectively it becomes very easy for video retrieval. These can be retrieved by using retrieval algorithms. After key frames are extracted the next step is an extraction of features, mostly these are extracted offline, the features are of two types: low level and high level. Low level features are object motion, color, shape, texture, loudness, power spectrum, bandwidth and pitch are extracted directly from video and stored in the database. High level features are semantic features like timbre, rhythm, instruments, and events involve different degrees of semantic features.

In this paper, the main objective of the proposed system is to show most specific videos regarding healthcare for a given user search query. Healthcare videos are easily available from hospitals or clinics. After the collection of healthcare video data, speech to text conversion is performed to create transcription snippets. Prescription reports are used to tag videos while uploading the video. Finally, keyword based labeling is done with the help of transcription snippet and doctor's prescription data. After getting video search results for a given search query, a user can give relevance feedback about the video search results. This relevance feedback is useful to refine the search results in the form of categorizing video search results into relevant and non-relevant groups. These groups are labeled with keywords, which in turn help to expand search query.

The rest of the paper is organized as follows: Section II contains the literature survey about related work. Section III contains the description of the proposed system. Finally, paper is concluded in the Section IV.

2. Literature Survey

B. Patel and B. Meshram [1] in 2012 proposed CBVIR (Content Based video indexing and retrieval in which they proposed that video indexing was a process in which videos were tagged and organized in a refined manner to get faster results. They used features like color, texture, shape, motion, object, face, audio, and video. More number of features they used more correct results they got. The color descriptors were also widely used, these were divided into two categories: global and local. The global color descriptor specified the overall calorie content, but with no information on the spatial distribution, the local color descriptor overcame this issue. Most prominent color descriptor used was MPEG-7. The audio features used by them were: short time energy, the main advantage of this is that it separates speech from non-speech in a given video which in turn helped in differentiating noise, as noise have short interval of time. Pitch used to find the speaker's excitement or tone. Pause rate was used to check the time of pause per frame the speaker used. Onset detection, which was based on the Hanning analysis window was used to find mid – level representations that aimed at localizing transients in a video. Shape Features were used as there is a statistical pattern recognition approach that is used for many years. There were issues like query language design, indexing of high dimensional frames, DBMS issue for video retrieval.

A. Bhute and B.Meshram [2] had done a review on text based approach for indexing and retrieval of image and video in which he said that text extraction from video is a process and has following steps, detection, localization, tracking, extraction, enhancement, and finally recognition of the text. The text localization can be again classified into two: texture based and second, region based, which are further divided into two: connected component (CC) and edge-based, these are like flow charts and work in a bottom-up manner and text are marked by rectangles. There are many theories proposed on connected components, the CC based methods have four stages:

I. Preprocessing i.e., noise reduction and colour correction.
II. Generation of connected component.
III. Non-text components are filtered out.
IV. Last is the grouping all the components.

Binarization techniques are the simple method which uses global or local or adaptive thresholding processes. Lienhart, use block matching algorithm, an international standard for video compression and also used temporal text motion to clear the extracted text regions. For a

specified block the algorithm was applied and desired results were getting.

R. Patil and C. Nayak [3] has proposed a system, pacify based video retrieval system, in which they used following a. Techniques in video data management: i. Video Parsing, in this the video is broken into key frames. ii. Video indexing: information is retrieved about the frame for indexing from the database. iii. Video retrieval and browsing: the users use queries to retrieve the video. The video is converted into number of scenes, the scenes are then converted into number of shots, the shots are then converted into the number of key frames, then by using OCR and ASR methods the text is retrieved, b. Retrieval of video using speech recognition: this is a process of capturing and converting any sound from microphone and converting it to a set of words, the set of words can independently have meanings or command, it has the following steps: i. Signal processing, ii. Speech recognition, iii. Semantic interpretation, iv. Dialog Management v. Response generation, vi. Speech synthesis. The method used by them was if a user or speaker speaks anything as an input to the system, the spoken content is searched and accordingly the results are getting and if the user types in a query, the text content is searched and the results are getting, they even used a combination of both the methods.

C. Jawahar, B. Chennupati, B. Paluri, et.al [4] proposed a system by which video can be retrieved based on text queries, they used to extract the text from videos and used them for faster retrieval of videos. The text within the videos can be indexed and has relation with the semantic information, this basically works on OCR (Optical Communication), the first step here was to extract text blocks from a video. The textual regions are used for indexing and retrieval of video. This method was dependent on two main aspects,

I. Quality of the video II. OCR availability

There are two types online and offline phase by which the conceptual diagram of video retrieval can be made. In offline mode, videos from various languages are converted into frames, a part of the important text region containing whole information about the videos are stored in the database and from this information video is retrieved. In online mode a query is entered from graphical user interface, the query gets converted into an image and the same process is followed as in offline mode. The results are based on a ranking system. As soon as a query is entered by the user it gets converted into key frames or key objects which are compared to the key words or objects present in the database and video is retrieved. This method has several drawbacks as there may be many videos for the same keyword and the results may not be correct.

D. Patil [5] did a survey on content based lecture video retrieval, in which she found that video can be searched or found from repositories in two ways: I. Metadata-based and II. Content based. Video title, video description, user feedback or comments come under metadata based, but the problem is that we don't know where the keywords exactly come in the video, in lecture videos there are three contents I. Lecturer speech: part of the video which shows speaker, speaking, II. Slides: part of the video which shows the slides, III. Lecture notes: part of the video which shows the board on which the speaker writes. The content based uses metadata to extract key frames to create indexes, this process is manually hard to do. NPTEL, MIT Open Courseware are some of the existing lecture video repositories or even YouTube. Tuna et.al. was the first to present their approach for lecture video indexing and search. They converted a video into sections and then the sections were converted into shots and then using OCR they converted the shots into key frames separating the speech and text to form a metadata and save the text file, which contained all data in the database, this text file was then used to search the relevant video from the database.

W. Bailer, H. Mayer, H. Neuschmied and W. Haas [6] proposed content based video retrieval and summarization using MPEG-7, in which they used IMB (Intelligent multimedia library), which helps in audio retrieval and semantic video. There are three main components of IMB: I. Annotation and Content analysis component, II. Multimedia database, III. Retrieval component. The multimedia objects with MPEG-7 together are send to the multimedia database. The multimedia objects have general metadata (text annotation and shots) and content based metadata (text from videos). In multimedia retrieval there was a web server, which used to forward the queries to the database and there the queries are grouped and the results were displayed on graphical user interface.

S. Mohamadzadeh and H. Farsi [7] came up with content based video retrieval based on HDWT and sparse representation. In this the video is converted in shots using CSS (Candidate Segment Selection), the main usage of this was to remove the complexity of non-boundary frames. Then they used Cut transition detection method in which the candidate, CT segments were

extracted using normalized hue-saturation-value (HSV) colour histograms, the shots were extracted using the shot boundary algorithm, after this the frame were extracted by using unsupervised clustering method. Video motion is classified into two: foreground motion and background motion, due to these motions, it was easy to convert the shots into key frames. The key frames contained text files in which the information about the video was present and it was used for retrieval of the video.

Relevance feedback [9] is a mechanism through which a video retrieval system generates a set of results for a given query, and then the user is allowed to send feedback in terms of relevant or non-relevant to the video retrieval system to improve search accuracy or to show more relevant video search results. This provides an interactive way for users to refine the retrieval results. This technique requires user input to identify positive results by labeling those which are relevant to the query. The relevance feedback information needs to be included with the original query to improve retrieval performance, such as the well-known Rocchio Algorithm [8].

3. System Development

3.1 Proposed System Architecture

In below section, the most important operations involved in this approach in order to search relevant videos which matches the user's query or intent are specified. The flow of the proposed scheme is depicted in Figure 3.1.

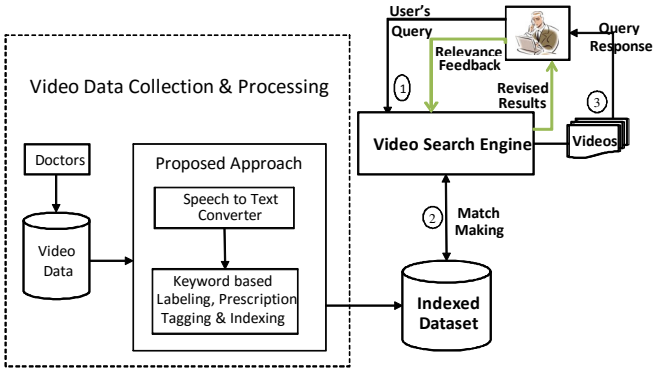

Fig 3.1 Proposed System Design Flow

3.2 Brief Description

Given approach is inspired by the video information available in the Healthcare Clinics. It explores the relevant videos to carry out patient's effective diagnosis and these videos are also informational to newer practitioners at the healthcare clinics. These videos describe and let out the diagnosis and treatment carried out for a particular patient suffering from a peculiar disease. These videos related to a particular disease are grouped into one cluster and less relevant into another cluster. In the proposed system, speech to text conversion along with the prescription data is used to show more significant videos for a given search query. Videos are labeled with the help of transcription text and prescription tags. These labeled keywords assists video search engine to discover the relevant videos with greater precision.

Figure 3.1 illustrates the architecture of the proposed system to add a video search engine for the healthcare data. A search engine's uses the doctor and patient's communication videos while disease diagnoses and apply speech to text conversion in order to label the videos. When a user requests the video search engine for a desirable objective (step 1), it uses preprocessed video data, i.e. healthcare data to match the query or intent (step 2) and return the most specific video (step 3) that satisfy the requested objective. Finally the user can submit the relevance feedback about the video search results displayed. This helps to rearrange the search results as well as query expansion.

3.3 Healthcare Data or Video Data

In the web search environment, users need is fulfilled with information relevant to the user's search query. Information can be of any type such as, videos, audios, images or any textual data. Among these videos are more informative and useful, as video data give audio-visual representation simultaneously. In this framework, doctors conversation videos recorded while communicating with patients are taken as video data. These videos are related to Healthcare. So this video data is also referred as Healthcare Data. Here, Healthcare data contain the videos related to various diseases and its diagnosis by the specialized doctors in respective streams.

These videos are explanatory and valuable to carry out effective diagnosis, if same patient is not cured with previous treatment. This helps to raise the success rate of the effective diagnosis process within a short period of

time, in case of highly viral or pathological diseases. Therefore, the given method utilizes healthcare data as a relevant decision in assessing patient disease successfully. As healthcare videos are easily available or can be recorded at low cost.

3.4 Speech to Text Conversion to label healthcare videos

As the video search engine shows the most relevant videos for a given user's query. In order to increase the search efficiency the videos speech is converted into text document i.e. transcription snippet. This transcription information can be used to label the videos with high frequency count words in the transcription document of a particular video. These high frequency count words are referred as keywords for that particular video. These keywords when used for labeling, i.e. naming the video then the search engine can show the effective results for the ambiguous query. The name of video is not enough to get the meaning of the query submitted. In order to get more information, each video is fed with additional text content through the keywords that is obtained from an individual video. Text pre-processing is done on transcription contents, converting words into lowercase, removing stop words (frequent words) and word stemming using porter stemmer algorithm.

Lastly, TF-IDF vector of videos transcription snippets are formed respectively as,

$$T_{ui} = [\, t_{w1}, t_{w2}, \ldots, \ldots, t_{wn}]^T \qquad (1)$$

Where, T_{ui} TF-IDF vector of videos textual snippet. U_i is i^{th} video in result. The t_{wj} denotes j^{th} term in the video's textual snippet respectively. Each video is represented by T_{ui} , this is nothing but keywords in the transcription snippet, which can be used for matching user search query. These transcription snippets contain user requirement, which in turn is used to refine a search query.

High frequency count keywords are extracted with the help of above equation number 1. These keywords and doctor's prescription tags assists keyword based video search engine to get effective search results for a given particular query.

Keywords can be used to represent the user search intention for a particular group or cluster, which in turn used to categorize the search results. The depicted keywords can be used to recommend more significant and precise query.

3.5 Indexing Video Dataset

The main purpose to store an index is to optimize speed and performance while finding relevant videos for a search query. Without an indexing a dataset, search engine takes considerable time and computing power to get the required content as it scans each and every data in the corpus. Video indexing is the process of providing viewers a means to approach and navigate contents easily and accurately. Full-text indexing is done with the videos in the dataset to quickly search for information in databases.

3.6 Restructuring Video Search Results

Video search results are restructured on the basis of user's relevance feedback for existing user search intents. User search intents are represented with most relevant keywords. Each video is categorized into a group by matching between user search query vectors and video's keyword vector.

4. Conclusions

The proposed system can be used to show most relevant video search results for a query by performing the match making. First, video data i.e. healthcare data collected and then the speech to text conversion is performed in order to give precise labels for each video. The preprocessed videos can be used as input for the specific video search engine. Using proposed system, user's relevance feedback is used to restructure video search results. Restructured video search results groups are labeled with keywords. This helps the user to get exact information as per users need in a more effective way. The discovered results group's labels can also be used to assist users in refining video search. The proposed approach considers the videos of the healthcare data. As these videos are shorter in time, so the running time is usually short. The complexity of the proposed approach is low and it can be used in reality easily. This proposed system can be utilized to track the patient's treatment, to impart an effective disease prescription and health diseases analysis.

Acknowledgments

Tanvir Ambekar is thankful to Dr. Vijaya B. Musande, Professor, H.O.D., Computer Science & Engineering

Department, Jawaharlal Nehru Engineering College, Aurangabad, for her constant support and helping out with the preparation of this paper. She is also thankful to the Principal, Jawaharlal Nehru Engineering College, Aurangabad for being a constant source of inspiration.

References

[1] B. Patel and B. Meshram , "Content Based Video Retrieval Systems", International Journal of Ubi Comp (IJU), Vol.3, No.2, April 2012.

[2] A. Bhute and B. Meshram, "Text Based Approach For Indexing And Retrieval Of Image And Video: A Review", Advances in Vision Computing: An International Journal (AVC) Vol.1, No.1, March 2014.

[3] R. Patil and C. Nayak ," Pacify based Video Retrieval System", International Journal for Rapid Research in Engineering Technology and Applied Science,February 2016.

[4] C. Jawahar, B. Chennupati, B. Paluri and N. Jammalamadaka," Video Retrieval Based on Textual Queries".

[5] D. Patil and M. Potey , "Survey of Content Based Lecture Video Retrieval", International Journal of Computer Trends and Technology (IJCTT) – Volume 19 Number 1 – Jan 2015.

[6] W. Bailer, H. Mayer , H. Neuschmied, W. Haas, M. Lux and W. Klieber, "Content-based Video Retrieval and Summarization using MPEG-7".

[7] S. Mohamadzadeh and H. Farsi, "Content Based Video Retrieval Based On Hdwt And Sparse Representation", Image Anal Stereol 2016;67-80 ,March 2016.

[8] J.J. Rocchio, "Relevance feedback in information retrieval", In The SMART Retrieval System-Experiments in Automatic Document Processing, pages 313-323. Prentice Hall Inc. 1971..

[9] Mark van Uden, "Rocchio: Relevance Feedback in Learning Classification Algorithms".

[10] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval.ACM Press, 1999.

[11] Porter, M. An algorithm for suffix stripping. Program, Vol. 14(3), pp. 130-137, 1980.

[12] Magdalini Eirinaki and Michalis Vazirgiannis, "Web mining for web personalization", ACM Transactions on Internet Technology, 03(01):1-27, February 2003.

Author -

Tanvir Ambekar received the BE degree from Dr. Babasaheb Ambedkar Marathawada University, Maharashtra, India, in 2011, and is pursing ME degree from Dr. Babasaheb Ambedkar Marathawada University, Maharashtra, India, both in Computer science and Engineering. Her current research interests include information retrieval and image processing. She is a student member of the ACM.

Vijaya Musnade BE and ME degrees from Dr. Babasaheb Ambedkar Marathawada University, Maharashtra, India, in 1995 and 2007, respectively, and the PhD degree from Dr. Babasaheb Ambedkar Marathawada University, Maharashtra, India, in 2012, all in Computer science and Engineering. Since 2010, she has been a professor and Head of Computer Science and Engineering Department, Jawaharlal Nehru Engineering College, Aurangabad. Her current research interests include remote sensing, image processing, machine learning, computer vision. She is a member of the IEEE.