Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 06, June -2018

SENTENCE FRAMING USING VIDEO

Aditya Ravindra Kale, Ankita Baburao Kate, Apurva Dinesh Kamaji, Om Arvind Kolte,
Prof. S. R. Kakde

Rajarshi Shahu College of Engineering (RSCOE)
Tathawade, Pune

Abstract: *To cope with the large amount of online videos we downloaded from YouTube, we transform every video shot into a set of images by extracting the video shot to interpret the videos for better understanding the content that it holds. Here we convert the video into query images and store it as testing dataset. In this phase, we make use of Gabor feature and wavelet transform on the query image cluster extracted from the query video shot V to extract the various features out of it. Visualization of the Gabor Filter for edge detection and extraction of texture for simple and complex cells. Wavelet transfer uses low pass filter and high pass filter to detect the various edges and textures. Extracting the features of the training image and these the features are stored in the array format. We extract features for the testing data set also. We further make use of machine learning to compare and find out perfect match of query image. The data set plays important role. In these step we making use of machine learning's neural network wherein we give testing data set as in input node and compared with the nodes to give the correct output. The best matched image is then compared with the various classes in order to extract the keyword respectevely. The keywords are then classified mainly as per the three domains namely object, event, scenes. We then frame a sentence using acquired keyword.*

I Introduction:

Contents of online videos are too complex to describe artificially, so let alone automatically generate a sentence. Second, it is not easy to accumulate videos as training data sets since most of the online videos have no labels. Meanwhile, it is time consuming to manually annotate a huge amount of videos with sentences. Last but not least, representing contents of videos with natural languages is more convoluted and multifaceted compared with independent tags: it needs not only to estimate objects in videos but also objects actions and scenes of events. What is more, the correct grammar is another factor we should consider of. WITH the prevalence of social multimedia in the 21st century, digital images have become more and more accessible to the public. As time goes by, however, simple images can no longer meet peoples demand and other more informative media are needed. Therefore, videos become increasingly popular. And technologies assisting users to search and understand contents of videos are required. Conventional approaches to video annotation predominantly focus on supervised identification of a limited set of concepts. However, many ambiguous meanings will be introduced when only the keywords are provided for searching the video. Therefore, studies have been conducted on annotating videos with sentences. Since video content includes objects, events, and scenes, generating sentences for videos will help to better understand the underlying activities happening in the video and there will be no ambiguity introduced. Moreover, if a video is annotated with a sentence, it is easier for users to search with flexible queries. While the idea of annotating videos with sentences is promising, there are several challenges. First of all, contents of online videos are too complex to de- scribe artificially, so let alone automatically generate a sentence. The generated frames from first step and then extracts the features using Gabor filter and Wavelet transform. In the third step, the extracted features are then compared with the features which are present in the training dataset. In the fourth step, if the features get a match with the features in training dataset, then keyword is generated with the help of Artificial intelligence and Neural Networking. In the fifth step, a proper sentence is formed with the help of keywords and appropriate use of prepositions and verbs. The proposed system will thus help a user to see and understand the contents of a video in real time, thus making the informative media more efficient.first step is video preprocessing i.e. conversion of video shot into multiple sequences of frames. From the generated frames, features are then extracted. The second step takes the generated frames from first step and then extracts the features using Gabor filter and Wavelet transform. In the third step, the extracted features are then compared with the features which are present in the training dataset. In the fourth step, if the features get a match with the features in training dataset, then keyword is generated with the help of Artificial intelligence and Neural Networking. The second step takes the Second, it is not easy to ccumulate videos as training data sets since most of the online videos have no labels. M anwhile, it is time consuming to annually annotate a huge amount of videos with sentences. Last but not least, representing contents of videos with natural languages is more convoluted and multifaceted compared with independent tags: it needs not only to estimate objects in videos but also objects actions and scenes of events. What is more, the correct grammar is another factor we should consider of.

Literature Survey

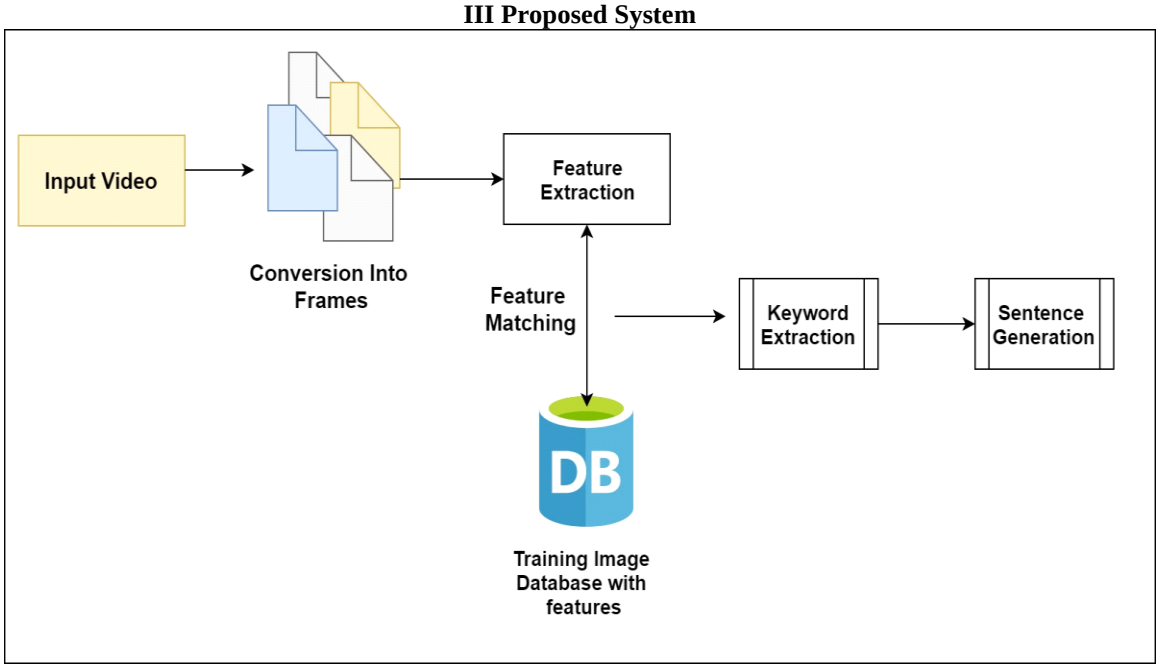
1. A S Wang[1] Video annotation (also widely known as video concept detection or high-level feature extraction), which aims to automatically assign descriptive concepts to video content, has received intensive research interests over the past few years. Various methods are put forward to automatically annotate videos with words. There are several works with respect to TRECVID , an annually video retrieval contest with the goal of creating a stock of best practice for video retrieval.

2. Wang et al [2] propose a learning based approach for video annotation. They learn the concepts in videos using the graph fusing the following factors: multiple modalities, multiple distance functions, and temporal consistency. Wang et al.[2] propose a novel semi supervised learning algorithm, named semi supervised learning by kernel density estimation, which is based on a nonparametric method, and therefore, the model assumption is avoided. Some work on video annotation only concentrate on one kind of videos like sports video, traffic video, or surveillance video.

3. P S John[3] propose an efficient method to annotate products in videosin. It collects a set of high-quality training data by mining information from Amazon and Google to build visual signature for each product. Then noise is removed by a correlative sparsification approach to refine the visual signatures that are used to annotate video frames.

4. Mossi et al [4] present a software application for annotating traffic videos with ground truth. Most of the methods on image or video annotation only generate nouns as their results. How to decide the verb, however, is still an obstacle in sentence making. Therefore, some previous works have focused on this aspect. At first, they obtain verb candidates by generating search queries for a given image with initial noun tags and establishing a sentence corpus from those queries. Then they further re-rank the candidate verbs with the tag context discovered. from the images both semantically and visually similar to the given image in the MIR Flickr data set. The above works aim at annotating images or videos with words.

III Proposed System



we have seen an approach for sentence generation from videos also good use of well-labeled image data sets to find sentence related elements. These elements include four main parts: object (the subject of the sentence), event (the action), scene (the place where the action happens), and adjective (modifier of the scene). This helps to gain a series of descriptive vocabularies for the video shot and generate a sentence to state the topic of the video. However, the corresponding video sentence generation approach can be further improved from following three aspects. First, the structure of our generated sentences is quite simple now.

IV Overview

With an increase in the use of informative media technologies and tools required to assist the uses to understand the contents of the video and the other informative media is of utmost importance. Also, due to the popularity of online video sharing websites such as YouTube, millions of users have treated online video as a source of information and entertainment. Hence, we need to find the solution to this uprising common problem. Therefore, video annotation has evoked great interest in the past few years. In present scenario, the content of the video is understood by subtitles if one is not able to understand the language of the video. But with the help of proposed system, a real time video annotation will be there which will be extremely helpful in emergency situations such as of accidents, burglary, etc. In this proposed system, a five-step approach to automatically annotate video shots with sentences is provided.

- The first step is video preprocessing i.e. conversion of video shot into multiple sequences of frames. From the generated frames, features are then extracted.
- The second step takes the generated frames from first step and then extracts the features using Gabor filter and Wavelet transform.
- In the third step, the extracted features are then compared with the features which are present in the training dataset.
- In the fourth step, if the features get a match with the features in training dataset, then keyword is generated with the help of Artificial intelligence and Neural Networking.
- In the fifth step, a proper sentence is formed with the help of keywords and appropriate use of prepositions and verbs.

The proposed system will thus help a user to see and understand the contents of a video in real time, thus making the informative media more efficient.

V Features:

Gabor filter

It is used to detect and pick up object or texture form the types of the different background The main thing is detected is the texture feature by filtering the frame vertical, horizontal, diagonal. The different parameters used such as wavelength, bandwidth, aspect ratio. Step used for the filtering and detecting if the wavelength is grater then bigger the strips length and if the wavelength is smaller then the smaller the strip length. Also can set the various angles for scanning the frame and detect the feature value. Second is bandwidth in this case larger the bandwidth, envelope increases thus the number of strips increases. Smaller the bandwidth number of strip also get decreases. The aspect ratio are used if the aspect ratio is high then only one pixel is selected and if the aspect ratio is small then more number of feature.In the first pass the detect vertical texture and outline and second pass horizontal texture and outline getting the detected entity. Also detecting the lighter and darker background the object trend to be dark and background to be dark.

VI Conclusion

In this report, we have seen an approach for sentence generation from videos also good use of well-labeled image data sets to find sentence related elements. These elements include four main parts: object (the subject of the sentence), event (the action), scene (the place where the action happens), and adjective (modifier of the scene). This helps to gain a series of descriptive vocabularies for the video shot and generate a sentence to state the topic of the video. However, the corresponding video sentence generation approach can be further improved from following three aspects. First, the structure of our generated sentences is quite simple now. We will dig deep on how to produce more complex sentences by adding more sentence elements and modifiers. What is more, we have not taken the problem of multiple objects into consideration yet. At last, the scale of our image data sets is relatively small and the number of all our elements is far from adequate. We will collect more images and elements in order to explain more videos.

VII References

1. Xueming Qian, Member, IEEE, Xiaoxiao Liu, Xiang Ma, Dan Lu, and Chenyang Xu ”What Is Happening in the Video? Annotate Video by Sentence ”

2. B. David G. Lowe Computer Science Department University of British Columbia Vancouver, B.C., Canada lowe@cs.ubc.ca January 5, 2004 ”Distinctive Image Features from Scale-Invariant Keypoints ”

3. C. Anish Talwar, Yogesh Kumar UG, CSE Dronacharya Group of Institutions, Greater Noida talwar.anish@yahoo.in *2UG, CSE Dronacharya Group of Institutions, Greater Noida "Machine Learning: An artificial intelligence methodology "

4. D. X. Qian et al., "HWVP: Hierarchical wavelet packet descriptors and their applications in scene categorization and semantic concept retrieval, " Multimedia Tools Appl., vol. 69, no. 3, pp. 897-920, 2014.

5. E. A. Altadmri and A. Ahmed, "A framework for automatic semantic video annotation: Utilizing similarity and commonsense knowledge bases, " Multimedia Tools Appl., vol. 72, no. 2, pp. 1167-1191, Mar. 2013.

6. F. Farhadi et al., "Every picture tells a story: Generating sentences from images, " in Proc. 11th Eur. Conf. Comput. Vis. Part IV, vol. 21. 2010, pp. 15-29.

7. G. Sally Goldman; Yan Zhou, "Enhancing Supervised Learning with Unlabeled Data ";, Department of Computer Science, Washington University, St.Louis, MO 63130 USA.

8. H. Zoubin Ghahramani, "Unsupervised Learning ",Gatsby Computational Neuroscience Unit, University College London, UK.

9. I. Rich Caruana; Alexandru Niculescu- Mizil, ";An Empirical Comparison of Supervised Learning Algorithms ";, Department of Computer Science, Cornell University, Ithaca, NY 14853 USA

