DOI:10.22413/ijatest/2021/v6/i4/5

INDEXING AND RETRIEVAL SYSTEM FOR SPEECH ANNOTATED DIGITAL IMAGES

[1]B.ABDUL RAHEEM , [2]G.SREEVALLI, [3]P.PREETHI , [4]M.PREMKUMAR REDDY , [5]V.ANIL

[1]Professor, Dept. of ECE,

[2,3,4,5]Graduate Scholar, Dept. of ECE,

Annamacharya Institute of Technology and Sciences,Rajampet, Andhra Pradesh, India.

Abstract: ***The abundant growth in use of digital camera and mobiles for picture has created huge storage requirements. The problem can be solved by constantly transferring it to hard drive of a computer or cloud. But indexing and retrieval system for digital pictures has become a known and vital problem. Most recently, automatic voice recognition technology, speech gloss and retrieval have given an option to retrieve them. Also predictable strategy for existing photograph ordering, recovery methods and repetitive work of manual writing is supplanted. Image annotation to enhance web photo search and image control is most preferred nowadays. In speech annotation and retrieval, a hybrid mechanism is utilized to assimilate picture like styles, syllables, words and characters. In the proposed system, speech annotated technique is used to provide the appropriate image for more number of search results. This technique has two processes to perform: firstly, offline process where words are trained using volume package and these trained words are stored in database, secondly, online process, content based refining is done on color, shape and texture.***

Keywords: ***Image Retrieval, Content Based Image Retrieval, Speech to text conversion, Bag-of-Features***.

I. INTRODUCTION

The Most usual and common methods of image retrieval in computer systems exploit various methods of adding metadata such as captioning, keywords, title or descriptions to the images so that retrieval can be performed over the annotation words. Image annotation, manually is very time consuming, laborious and expensive to address, extensive research has been done on automatic image annotation techniques [1]. The increase in use of social-web applications and the semantic-web have motivated the growth of several web-based image annotation tools in the recent past. A typical general image retrieval system is shown in figure 1 below.

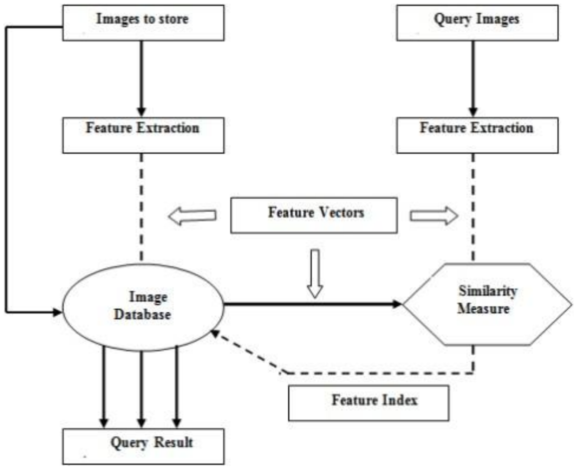


Fig.1. General Image Retrieval System

Image searching has been a specific type of data search used to get images from the database. For the search of images, the user provides query terms such as keyword, image file/link, or click on some image, and the system will return similar database images. The similarity between the query image and the database image could be Meta tags, distribution of colors in images, regions or shape details, etc. The search of images based on metadata associated with it such as keywords, text, etc. is nothing but image Meta search technique. In content based image retrieval system this is avoided, instead it retrieves the images based on similarities in their contents such as textures, colors, shapes etc. Novel exploration paradigms have evolved called image collection exploration. The image collection exploration is nothing but to explore large digital image repositories. It is a known fact that every day from different devices, huge amount of digital data in the form of images is produced; this brings challenges in storing, indexing and accessing from these repositories [2].

The metadata, (i.e, data that provides information about other data) of the image is stored in a large database and indexed accordingly, when a query to search for an image is requested, the search engine sees in the index, and the stored information is matched with the query. The results are obtained and presented in the order of relevancy. The efficacy of an image search engine depends on the results it returns, more relevant it is the best search engine and the ranking is done by ranking algorithms that are used in search engines.

Some search engines are designed to identify automatically some of the visual content, like, color, background, shapes etc. This is very much relevantly used in content based image retrieval system. When a search is performed, some thumbnail images are received and are sorted in the order of relevancy. Each thumbnail is linked to the website or

DOI:10.22413/ijatest/2021/v6/i4/5

repository. Some provide advanced search options linked to color, animations and sizes to receive more relevant image looking for [3].

II. EXISTING METHOD

The Content based image retrieval (CBIR) is a technique used to automatically retrieve images from a large database which are perfectly matched with the query image [4]. It is very hard to search the desired image from a large collection of images in a repository. Thus to search the desired image automatically, the need for development of an expert technique is desired. Primarily, two firm approaches are followed to retrieve the image from large database: text based search and visual based search. Generally the images are searched by text may be title (ex: in Google, yahoo, etc. searches) i.e., the images stored in text annotation and user types in a sequence of text to retrieve.

To enhance the precision and efficiency of the image retrieval system the content based image retrieval was introduced. The method uses visual content of the query image perfectly matched with the database images and deals similarities based on color, texture and shape. On the basis of similarity index the retrieved images are presented in order.

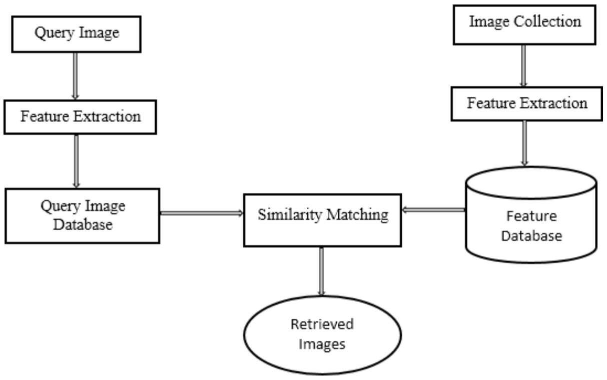


Fig.2. Block diagram of Existing method

The content of the image in terms of color, texture and shape is the heart of content based image retrieval system. The Figure 2 above shows structure of CBIR system. The similarity between the query image and database image, results in best matching between them. Every image stored in database consists of color, texture, shape features also some low level and high level features. Actually, these features are used to build the database in order. The image retrieval based on low level features can be obtained directly from the image whereas it is hard to solve arithmetic in high level features. Three primal techniques are used for image retrieval system: Image retrieval by Color, by Texture and by Shape.

A. **Image Retrieval by Color**: The color feature is a sensitive and most important feature of the image and the color histogram is used to represent this feature. The main advantage of color histogram is, it is fast and requires less memory. Also not susceptible to the changes in sizes and other parameters of image. On the basis of Hue, Saturation, Value, the feature vector of query image and database image is calculated. Feature extraction, histogram calculation, similarity and dissimilarity calculations, and arranging the images in order are the steps followed in retrieval of image by color features [5][6].

B. **Image Retrieval by Texture**: The Texture of an image is used for illustration constituents of images which are set uniformly. By texture segmentation the region of an image can be determined. The bounding boxes of the determine region of image may be used to retrieve the formation like an R-tree. The texture is effected by cross correlation and dimensions of the image and this can be solved by comparable techniques. The merit of texture based image retrieval technique is periodicity and scale. This is expressed in terms of contrast, direction and thickness. The accessing of related image by texture based retrieval technique involves two significant issues, the arithmetic analysis and the structural analysis. When the texture component is clearly identified, the structural analysis is performed whereas for micro textures the arithmetic analysis is preferred [7].

C. **Image Retrieval by shape**: Shape can be represented in two categories: the Region based and the boundary based. In the region based method the shape area of the image relating the inner quality i.e pixel quality is considered whereas in boundary based representation the external edge of the shape is considered. Here the outer description of the region, like the pixels near the object edge is measured [8].

The drawbacks of the content based Image retrieval are large dimensions of color images and hence the computations of such images are quite time consuming. They are sensitive to noise interference as illumination by color images is a matter of concern [9]. This system cannot handle rotation and translation, thus no support to speech annotations.

III. PROPOSED METHOD

The indexing and retrieval system for speech annotated digital images is carried out in two phases. Information based picture recovery and information based visual recovery. The former is the software of a computer vision technique bothers picture recovery, this is the problem of looking for virtual pictures in large collection of records [10]. The word content might confer to colors, shapes, textures, or other facts that may be a resultant form of picture. Without the potential to study the picture information, analysis must rely on metadata which may be difficult to create and also costly. The architecture for speech annotated image in CBIR is shown in figure 3 below. This process is carried out in two phases: firstly, converting speech to text by using python programming and secondly, taking this as input to retrieve the image with the help of MATLAB software.

DOI:10.22413/ijatest/2021/v6/i4/5

Speech recognition, researched at Bell Labs in the early days, having single speaker and used limited vocabularies, some dozens of words. Modern systems have much more capacity compared to the one used in early days. They can recognize a particular speech from multiple speakers and the vocabularies in different languages.

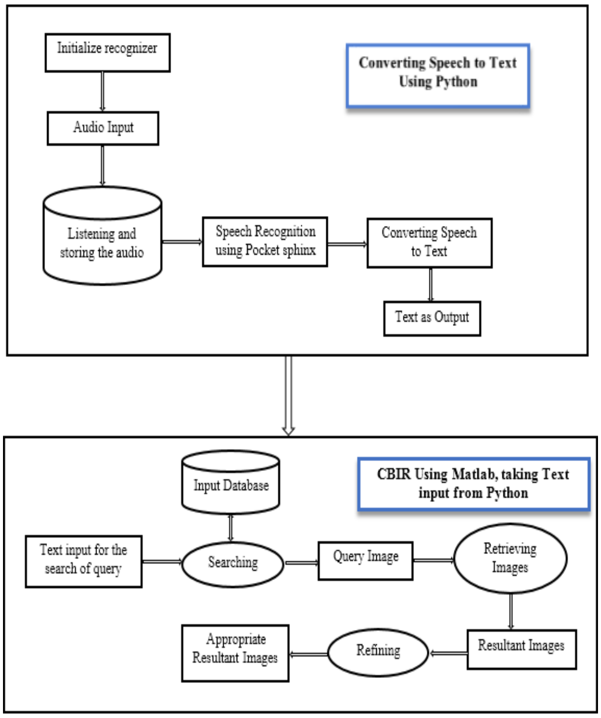

Initialize recognizer
Converting Speech to Text Using Python
Audio Input
Listening and storing the audio
Speech Recognition using Pocket sphinx
Converting Speech to Text
Text as Output
Input Database
CBIR Using Matlab, taking Text input from Python
Text input for the search of query
Searching
Query Image
Retrieving Images
Appropriate Resultant Images
Refining
Resultant Images

Fig.3. Block diagram of Proposed Method

Speech being the first component, must be converted from physical sound to an electrical signal by microphone, of course, and then by using analog to digital converter it is digitized. Once digitized, numerous models can be used to transcribe the audio to text. Most modern speech recognition systems rely on Hidden Markov Models, in which the statistical properties are assumed to be stationary for processing, generating group of vectors. To decode the speech into text, groups of vectors are matched to one or more phonemes. This calculation requires training and the whole process is computational extensive. Neural networks may be applied to simplify the feature extraction [11].

Secondly, the retrieval of images using customized bag-of-features, which is implemented on MATLAB. That is bag-of-features uses image features as the visual words that describe the image instead of actual words as in document retrieval [12][13]. The query's feature vector and feature vector of every image in the database is compared, usually using the cosine similarity. The refined appropriate resultant images are displayed, which is closest to the query image.

IV. RESULTS AND DISCUSSION

The process of converting spoken words into texts is known as Speech to Text conversion. Often it is called as speech recognition system, sometimes used to describe the process of extracting meaning from the speech, that is, speech understanding. The term can be avoided as it is the process of identifying the voice of a person: speaker recognition.

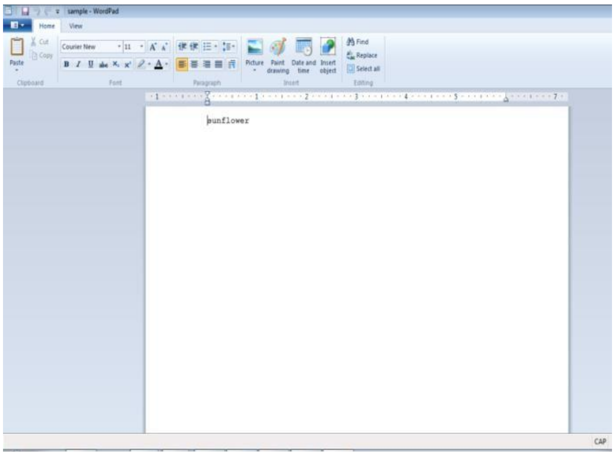
All speech-to text systems rely on atleast two models; an acoustic model and a language model. In addition, large vocabulary systems use a pronunciation model. However, there is no universal speech recognizer in the field. Thus the best transcription quality can be obtained if all the above said models can be specialized for a particular language or dialect, or type of speech and communication channel.

The speech recognition cannot be error free as the accuracy highly dependent on the speaker, style of speech and also the environment. Thus care has to be taken to convert and all the noise components that can effect have to be addressed before it is converted to text form.

The speech to text conversion is performed by pocket sphinx from python. Pocket sphinx is a part of CMU Sphinx open source toolkit for speech recognition. This package provides a python interface to CMU sphinx base and pocket sphinx libraries created with SWIG and setup tools. The recognizer class requires an audio data argument and it may be provided from the audio file or audio recorded by a microphone. The pocket sphinx recognizer converts audio into text and displays the output in the text format, which may be stored in a file of a specific format. The algorithm to convert the speech to text is shown below.

Algorithm for converting speech to text:

Step1: Initialize recognizer
Step2: Audio as input
Step3: Listening and storing the audio
Step4: Speech Recognition using Pocket sphinx
Step5: Converting Speech to Text
Step6: Text as Output



sunflower

Fig.4. Speech to Text Converted output

The second phase of the process is retrieval of images using customized bag-of-features, which is implemented using MATLAB. The bag-of-features is a technique adopted to image retrieval from the world of document retrieval. Here visual words that describe the image features is used instead of actual words as in document retrieval. The image features are used to measure the similarity between images such as color, texture, and shape. The advantage is can create visual word vocabulary customized to fit the application. For small database of images brute force search can be performed, whereas for large dataset bag-of-features is preferred because it provides a concise coding scheme to represent a large collection of images using a sparse set of visual word histograms. This enables compact storage and efficient search through an inverted index data structure. The algorithm for the above process is given below. In the example here, the steps to create image retrieval system for searching a flower dataset having 1000 images of different types is processed.

Algorithm for converting text to resultant digital images:

```
Step1:Text input for the search of query
Step2: Select the image features for retrieval
Step3:create a bag of features for the Query image
Step4:index the images for Retrieving
Step5:search for similar images and  Refine the images
Step6: Appropriate resultant images will be displayed
```

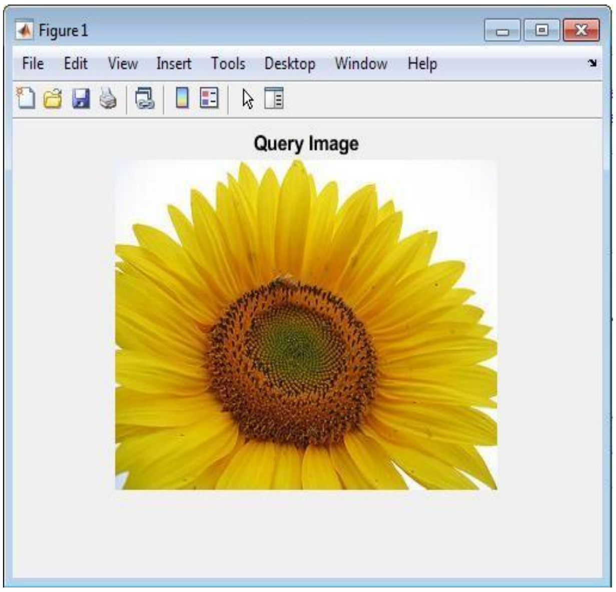


Fig.5. Query Image based on Text

The type of features used for retrieval depends on the type of images within the collection, based on that global image feature, such as color histogram or the local image feature extracted around the object key points is selected.

The extractorFcn is a function within the bagofFeatures class is used to extract color features, the function is described in the algorithm. With the feature type defined, the visual vocabulary is learned by using set of training images from bagofFeatures. Now the set of flower images can be indexed for search operation. Using the custom extractor function from previous step, the features are encoded into visual word histogram and added to image index. The final step is to use the retrieveImages function to search for similar images from the database. the retrieveImages returns the image IDs and the scores of each result sorted from best to worst as shown in figure 6.

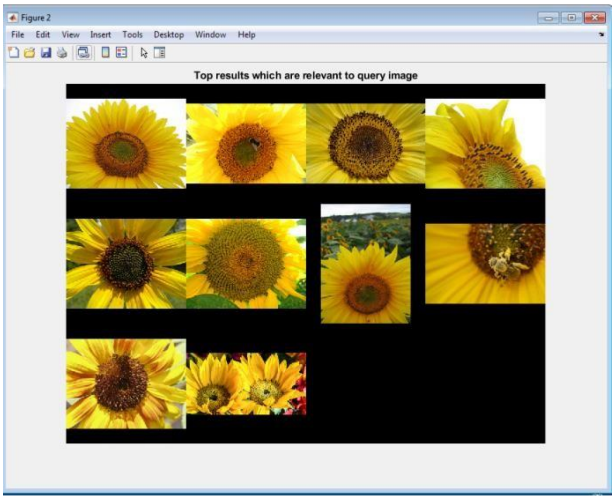


Figure 2
File Edit View Insert Tools Desktop Window Help
Top results which are relevant to query image

Fig.6. Output of Retrieved images based on Query Image

V. CONCLUSION

The paper presents a new image retrieval system for images, speech annotated in digital form, where in the voice or speech is converted to text and from then to extract the image from customized bag-of-features. The speech to text conversion takes place with the help of pocket sphinx recognizer on python platform then the query image is retrieved from the database by search algorithms executed in MATLAB. The results show that for large dataset this process is quite helpful although computation extensive. The process can be extended for more real time applications of robotics and embedded systems with neural network processes.

REFERENCES

[1] Ritendra Dutta, Dhiraj Joshi, Jia Li, and James Z. Wang, “Image Retrieval:Ideas, Influences, and Trends of the New Age”, ACM Computing Surveys, Vol 40, No 2, pp 1-60, 2008. DOI:10.1145/1348246.1348248.

[2] Jorge E Camargo, Juan C Caicedo, and Fabio A Gonzalez, “A Kernel-based framework for image collection exploration”, Journal of Visual Languages & Computing, Vol 24, No 1, pp 53-57, 2013. DOI:10.1016/j.jvlc.2012.10.008.

[3] Manish K Shriwas, and V R Raut, “Content based image retrieval: A Past, Present, and new Feature Descriptor”, in IEEE International Conference on Circuit, Power and Computing Technologies (ICCPCT), 2015.

[4] M A M Shukran, M N Abdullah, and M S F M Yunus, “New Approach on the Techniques of Content based Image Retrieval (CBIR) using Color, Texture and Shape Features”, Journal of Material Science and Chemical Engineering, Vol 9, pp 51-57. https://doi.org/10.4236/msce.2021.91005.

[5] R Vijaya Arjunan and V Vijaya Kumar, “Image classification in CBIR systems with colour histogram features”, in International Conference on Advances in Recent Technologies in Communication and computing, pp 593-595. 2009. DOI: 10.1109/ARTCom.2009.233.

[6] M Venkata Dasu, V R Anitha, Fahimuddin Shaik, and B Abdul Rahim, “Feature Extraction of satellite images using Decorrelation stretching and color scatter plots”, journal of Digital image processing, vol 3, No 14, pp 873-877, 2011.

[7] F Tomita, and T Saburo, “Computer Analysis of Visual Textures”, Kluwer, 1990. https://doi.org/10.1007/978-1-4613-1553-7

DOI:10.22413/ijatest/2021/v6/i4/5

[8] Roger Weber and Michael Mlivoncic, "Efficient Region Based Image Retrieval", ACM Conference on Information and Knowledge Management, USA. 2003.

[9] G Chandrasekhar, B Abdul Rahim, Fahimuddin Shaik, and K Soundara Rajan," Ricean code based compression method for Bayer CFA images", Recent advances in Space Technology services and climate change 2010 (RSTS & CC-2010), pp 102-106, 2010. DOI:10.1109/RSTSCC.2010.5712810.

[10] P Sreevani, B Abdul Rahim, and S Fahimuddin, "Spectral_spatial classification for Hyper Spectral Images based on an Effective Extended Random walker", International Journal of Engineering and Computer Science, Vol 4, No.7, pp 13110-13114, 2015.

[11] V Anand and B Abdul Rahim, " Speaker Identification using TESPER Technique & Neural Networks", in the Proceedings of National Conference on Advances in Communication Technologies, India. 2009.

[12] Sivic J and Zisserman A, "Video Google:A text retrieval approach to object matching in videos", Proceedings of Ninth International Conference on Computer Vision, pp 1470-1477, Vol 2, 2003. DOI: 10.1109/ICCV.2003.1238663.

[13] Philbin James, O Chum, M Isard, Josef Sivic and Andrew Zisserman, "Object retrieval with large vocabularies and fast spatial matching", IEEE conference on Computer Vision and Pattern Recognition, pp 1-7, 2007.
DOI: 10.1109/CVPR.2007.383172.