IJCSN International Journal of Computer Science and Network, Volume 4, Issue 2, April 2015
ISSN (Online): 2277-5420 www.IJCSN.org
Impact Factor: 0.417

Multimodal Visual Search

[1] Sushma Vanam ,[2] Deepali Shinde , [3] Ruchika Singh

[1] Department of Computer Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

[2] Department of Computer Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

[3] Department of Computer Engineering, Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract - This papers describes an interactive search system for mobile devices in which the input can be given through text, speech and image and output is obtained in the form of images. This eases the search that takes a long while to process by the way. This application is the efficient search technique for images that do not have a specific location or the image is not available with us. This image can be searched by entering the voice input. Similarly the text search is also implemented in the system that recognizes the search via text input ant accordingly specifies the result. In this paper, we propose a multimodal image search system that fully utilizes multimodal and multi-touch functionalities of smart phones. This system shows various input techniques and corresponding outputs according to the search. The cosine similarity, keyword stemming and threshold frequency determination algorithms are used to provide a more simplified result.

Keywords - *Image Retrieval, Mobile Search System, Voice Based Search Engine, Voice Search System.*

1. Introduction

The system implements various types of inputs to provide the user with ease of specification for their search. Thus, Voice based search engine is the applied system that provide the facility to user to get the required output in the most simplified form as it requires its output to be. This system is new approach to the previous search engines that generally take a single command as an input and provides the appropriate output. Most of the image system services use text query, but it's difficult for user to convert their search intent to text input.

Typing a text query, is a difficult for user on phone as phone has limited display and they can do spelling mistakes so it's difficult to system to search it. So other services introduced voice as a query that provides better interaction in search process. There are two aspect of web surfing - a) surfing the content of a page, b) navigation through these pages. Here, user have to creates his account to sign in and search the query so, he can views a page , that he have to search. The user can then look for his/her area of interest and then read the particular content or may decide to leave that page. He/she may choose to skip any amount of information in between to reach the desired area of interest. Among the early attempts, voice input present the web page in an easy-to-use interface and convert speech to text, having the different gender voice for reading texts and links.

Compare with text search, ma search, and photo-search, visual(image and video) search is still not very popular on the phone, though image search has become a common tool on the PC since 10 years ago, with which the user can input text query to retrieve relative images.

2. Literature Survey

Recently, many systems are developed that used multimodal query for image retrieval. Quickset was the first system that applied multimodal interaction with mobile systems, developed by the US Marine Corps. The another example is Speak4it local search application, where users multimodal commands that combines speech and drawing. Sometimes when a user performs searching, example images will not be always at hand, which motivates sketch based image retrieval (SBIR) research that uses simpler hand-drawn sketches as a query image. Sketch-based search is

more accurate and convenient .For example, if you want to find some picture of a beautiful pendant that you once saw in shop. That query is too complicated to search but the sketch of pendent is simple. User have to express their visual intent through sketches but it's difficult for the users without drawing ability. Yang Cao proposed Mind Finder system, which is the first interactive sketch-based multimodal image search engine. It provides users to sketch major curves of the target image in their mind; Tagging and clearing operation are also added for higher search results.

An image raw curve-based algorithm applied to calculate the similarity between the salient curve representation of natural images and a user's sketch query. The different visual search application applied different types of image matching techniques. Directly applying text based search to mobile visual search is straightforward. Traditional text-based search engines like Google and Bing are still available on mobile devices. However, lengthy text queries are neither user-friendly on phone, nor machine-friendly for search engine. The fact is that mobile users use only 2.6 terms on average for search, which can hardly express their search intent. As recent mobiles supports multimodal input such as the built-in high resolution camera, voice, multi-touch function , etc.., the user's search intent could be extended to more convenient and various measures than ever before .

Table I summarizes the recently development in mobile phones applications for multimodal search. As the speech recognition become mature, phone application is Apple Siri , which combines speech recognition, natural language understanding and knowledge based searching techniques. The user is enabled to make a speech conversation with the phone and get information and knowledge from it. It provides unprecedented ask-and-answer user experiments. The user can ask the phone for anything by only speech and get multimedia answers.

3. Proposed System

This paper proposed a new multimodal based search engine for image retrieval that helpful in two different situations. Consider an example a user move to an unfamiliar place and visits an unknown location .To visit the same location next time. 1) He simply takes a picture of that one. 2) Another situation is that he forgets to take pictures of place he visited, but can describe its particular appearance such as "a grass below the sky". The proposed system handles these two situations. In first case system uses an input such as a captured photo of the place and start searching process and retrieve similar images from the web. In the second case , user's doesn't have the existing image, but the user can generate an image query by giving speech input to the system, that represent picture described in the users mind.

Earlier development sketch based search engine to express user's visual intent in the form of a sketch, but it's difficult for users without drawing experience. But our system helpful for all users they can simply express their information needs via speech input. It uses Google Speech recognition engine services to convert user speech to text input. Then keywords extracted from the text. Based on these keywords , users can start searching process, but text-based image retrieval does not give more satisfactory results. So the proposed system generates exemplary images corresponding to each keyword from the back-end search engine (i.e. Google). The images results depends on the position and resize of exemplary images in the composite query decided by the user. To improve search results, location information also provided to the user. The architecture of the system is as shown in fig.1

Similarly, when a text is given as an input to the system, then this input gets processed via text parser, and the words are stemmed (i.e. bought to their root form) and the tags are inserted. After inserting the tags, using Cosine Similarity algorithm, the outputs are ranked accordingly and then processed. The output having highest priority is displayed first and then the output with lower than that. And this is then displayed in the form of the image as output.

Now, in case of image as an input, the image is first blurred and then it is converted into HSV from RGB to get the detailed data about the pixels and color of the image. This conversion to HSV is then followed by Histogramming the Image and quantizing it to remove the disturbances from the image such that the actual picture of what is to be searched is obtained and that the search can be easily implemented. Then the image is normalized and then the output ranking is obtained via cosine similarity algorithm. Thus this will give the output in the form of image. Thus the proposed system is an efficient technique for search implementation.

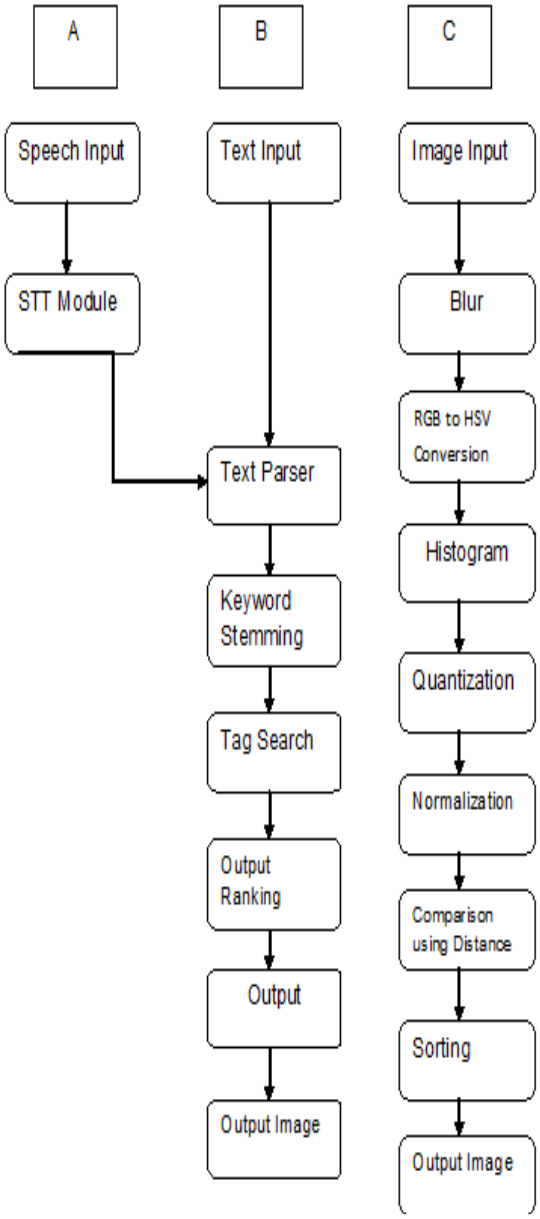

Fig.1 Architecture of the Proposed system

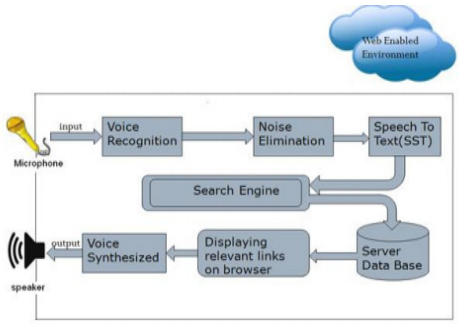

Fig.2 Proposed framework for voice input

4. Algorithms

The algorithms that are being used in the proposed system are

1. Cosine Similarity Algorithm
2. Keyword Stemming Algorithm
3. Image Processing Algorithms
   a. RGB to HSV Conversion
   b. Thresholding
   c. Edge Detection

The previous system used technologies such as JIGSAW+. The system, the Joint search with Image, Speech, And Word Plus (JIGSAW), takes full advantage of the multimodal input and natural user inter-actions of mobile devices. It is designed for users who already have pictures in their minds but have no precise descriptions or names to address them.

An input image sent to an existing search engine and finding matching images. 2) In some cases, the user doesn't have an existing image and have the picture description in their mind, then they prefer to use voice query to initiate search process. Image search using voice involves 4 major components (1) Speech recognition, (2) Keywords extraction, (3) Interactive exemplary visual query composition. Speech recognition is a difficult process than image recognition. It requires 90% accuracy environment. In our application, the Google speech recognition service converts speech to text input after those keywords are extracted from text and forwarded to the Google image search service.

ALGORITHM 1- COSINE SILMILARITY ALGORITHM

- Similarity and Positions: Retrieve the similarity for each query inserted in the reverted file; estimate the positions by averaging positions of the matched words.
- Find maximum similarities
- Merge the similarities from different entities
- Obtain the most viewed entity according to the viewed and ranked articles and display.
- Combine the Merged Similarity and position consistency

ALGORITHM 2- KEYWORD STEMMING ALGORITHM

- Stemming is used for reducing inflected or

derived words to their stem, base or root form.

- It reduces the input form into its root form first
- Then it looks up for the unrecognized words in the **look-up table** and then process it while that table maintains the record of the word used.
- The **suffix-stripping algorithm** removes the suffixes and brings the word to its original root form.

ALGORITHM 3.a – RGB to HSV CORSION

- The image is first separated in the RGB form, to recognize color
- Then for conversion for detailed image processing, the RGB converts the image in HSV form by separating the bits and shifting the pixels accordingly.
- Following formula is used

New color = (V) || (S<<8) || (H<<16)

ALGORITHM 3.b- THRESHOLDING

- It is an essential concept related with image processing and machine vision.
- Thresholding is a conversion between a grey-level image and a bi-level image.

ALGORITHM 3.c- EDGE DETECTION TECHNIQUE

- Edge detection is the process of locating the edge pixels.
- Then an edge enhancement will increase the contrast between the edges and the background in such a way that edges become more visible.

5. Mathematical Model

5.1Problem Description

Let, S is the system such that

S= {U,St, R,Ip,Tp,St,Co, I*,C,Im,Or,Op, }

Where,

U= is a set of user.

St = Speech to text conversion.

R= Resources like Processors, RAM are monitored by web server.

Ip =is a set of inputs.

Ip= {SI,TI,II}

Here there are three types of input that are speech input , text input,
image input etc.

Tp=is a set of text parser which is useful for parsing.

St=keyword stemming for comparing words.

Ip*={SI*, Ti*, II*}

Ip* is desirable input which selected by user for searching.

C= is RGB to HSV conversion of image.

Im=set of image processing steps i.e Histogram , Quontization , Normalization etc.

Or=it is set of output rating which is used for taking the required and
essential result.

Op=it is a set of outputs .

Co=it is useful for tag search using cosine similarity.

5.2 Activities

Activity 1: User gives the input to the system. Let f (A) be function of user. Thus, f(A)->{SI,TI,II} Є I .

Activity2: Application does the speech to text conversion if the input is speech, Let f (Ts) be function of speech to text conversion.

Activity 3: Application does the text parsing if the input is speech or text. Let f (Tp) be the function of text parsing.

Activity 4: Application does the stemming of text. Let f (st) be the function of stemming text.

Activity 5: Application does the tag search using cosine similarity. Let f (Co) be the function of cosine similarity.

Activity 6: if input is image then Application does blur and converts RGB to HSV form. Let f(C) be the function of conversion.

Activity 7: if input is image then application does image processing steps like
Histogram, Quantization, and Normalization etc. Let f (Ip) be the function of image processing.

Activity 8: Application does the output rating, Let f(Or) be the output rating.

Activity 9: Application gives the output. Let f (Op) be the final output.

5.3 Venn Diagram

Activity 1: User gives the input to the system.

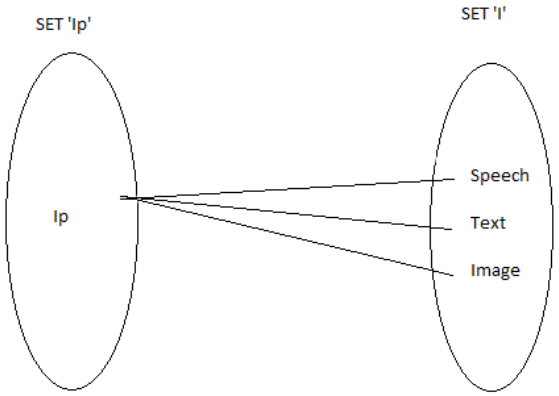

Fig 3: Activity 1

Activity 2: Application does the speech to text conversion if the input is speech

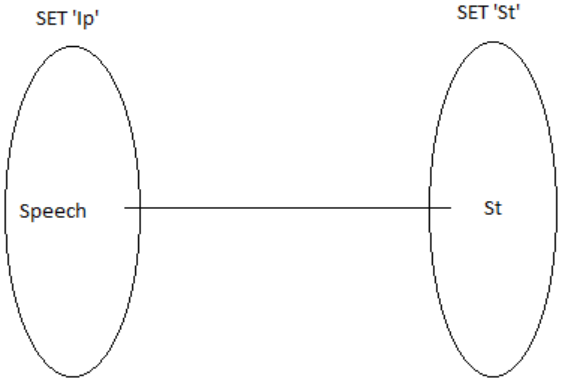

Fig 4: Activity 2

Activity 3: Application does the text parsing if the input is speech or text

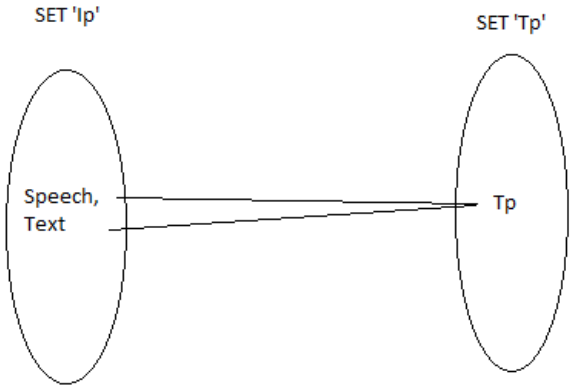

Fig 5: Activity 3

Activity 4: Application does the stemming of text

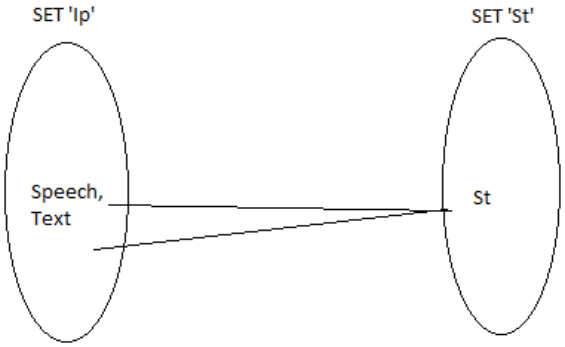

Fig 6: Activity 4

Activity 5: Application does the tag search using cosine similarity

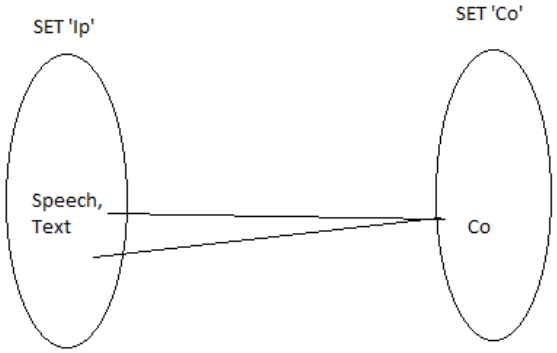

Fig 7: Activity 5

Activity 6: if input is image then Application does blur and converts RGB to HSV form.

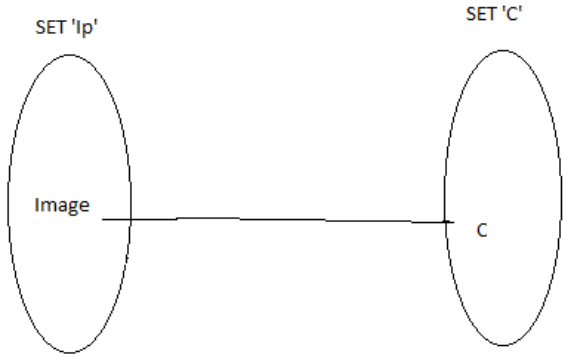

Fig 8: Activity 6

Activity 7: if input is image then application does image processing steps

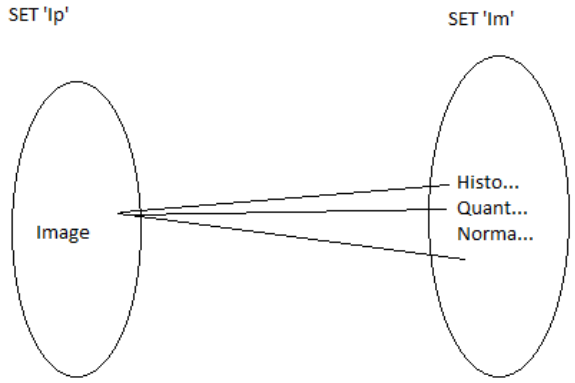

Fig 9: Activity 7

Activity 8: Application does the output rating

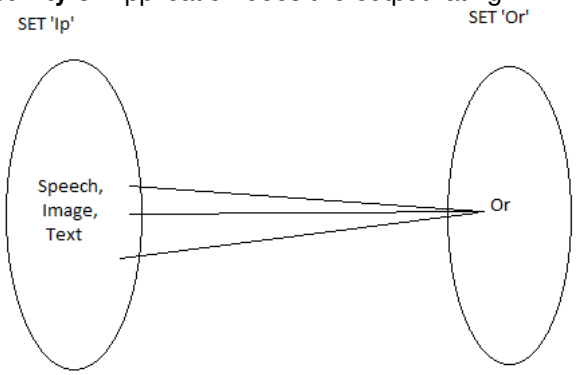

Fig 10: Activity 8

Activity 9: Application gives the output

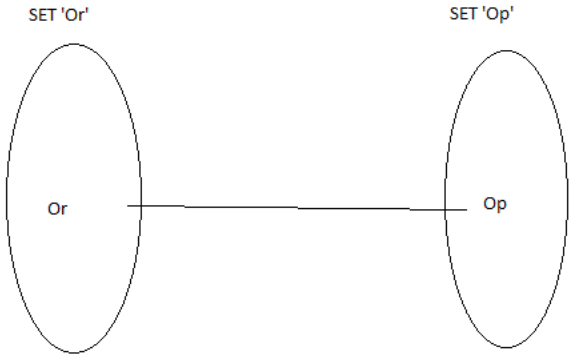

Fig 11: Activity 9

5. Conclusion

Here we propose a multimodal search system that helps users to express their needed information implicitly and explicitly. Only the user has to know what information they are looking for, so expressing their information plays a very important role in the search process. Proposed system gives a better way to express their query than other existing systems, it provides more relevant search results, especially in case where users can have a partial picture description in mind but the accurate information is required within a specified domain.. The System provides a cool game-like user interface for query formulation and enhanced user experience on mobile phones. It provides the accurate output in the form of the image so that the user gets what it demands for and not the other options to choose from. This will be an efficient search implementation.

6. Future Scope

The future scope of our project include the output in the form of voice when the input is voice, the output in text when input is in text form and the output in image form when the input is in image form. And the further scope of this projects include the inbuilt browser that will be contained or embedded in the application itself so that it won't have to search for any other browser in the device and connect to that browser to provide the output. The application will have an inbuilt browser and it will be a two way opener i.e. the application will not close once the search is implemented.

Acknowledgement

It is our great pleasure to express our deep sense of gratitude to Mrs. Pratiksha Dhande, Computer Science,

for her valuable guidance, inspiration and whole-hearted involvement during every stage of project preparation. Her experience, perception and through professional knowledge, being available beyond the stipulated period of time for all kind of guidance and supervision and ever-willing attitude to help, have greatly influenced the timely and successful completion of project preparation. We extend our sincere thanks to Mrs. Pratiksha Dhande, Project Guide for her valuable guidance. She was always there for suggestion and help in order to achieve this goal. We are indebted to Mrs. Deeksha Bharadwaj, HOD and Dr. R.D. Kharadkar, Principal, G.H. Raisoni Institute Of Engineering and Technology, Pune for encouragement and providing us the opportunity and facilities to carry out this work . And finally we would like to thank the college for being such strength during the entire work .

References

[1] International journal of Computer & Organization Trends – Volume 7 Number 1- Apr 2014 ISSN: 2249-2593

[2] Jianye Liu & Jiankun Yu, "Research on Development of Android Application," School of Information Yunnan University of Finance and Economics KunMing, China, IEEE research papers, 2011.

[3] Mihal Brumbulli, Blerina Topciu, Arbora Dalaci, "SMIS: AWeb-Based School Management Information System," Department of Computer Engineering, Faculty of Information Technologies, Polytechnic University of Tirana, Albania, 2008.

[4] Uduak A. Umoh, M.Sc.1, Enoch O. Nwachukwu, Phd.D., Eyoh, M.Sc.,"Object Oriented Database Management System: A UML Design Approach," University Of Uyo, Nigeria, Nov 2009.

[5] S. R.Bharamagoudar, Geeta, S.G. Totad, "Web Based Student Information Management System" Assistant Professor, Dept. of Electronics & Communication Engg. College Bagalkot Karnataka Associate professor, Department of IT,GMR Institute of Technology, RAJAM, Andhra Pradesh Professor , June 2013.

[6] Bongani T. Mabunda and Johnson O. Dehinbo Enhancing University Class Management System with Instant email feedback Alert", Oct 2012.

[7] Wei-Meng Lee, "Beginning Android Application Development", CrosspointBoulevard: Wiley Publishing. Inc. 10475, 2011.

Sushma Vanam pursuing graduate degree in the field Computer Engineering at G. H. Raisoni Institute of Engineering and Technology, Wagholi, Pune, under Savitribai Phule Pune University.

Deepali Shinde pursuing graduate degree in the field Computer Engineering at G. H. Raisoni Institute of Engineering and Technology, Wagholi, Pune, under Savitribai Phule Pune University.

Ruchika Singh pursuing graduate degree in the field Computer Engineering at G. H. Raisoni Institute of Engineering and Technology, Wagholi, Pune, under Savitribai Phule Pune University.