IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 3, March 2015.
www.ijiset.com
ISSN 2348 – 7968

Multimodal Image Retrieval on Mobile Devices

Pravin Kumar, Shailesh Kumar, Akshay Nimbalkar, Swapnil Pandey,

Computer Department
D.Y.Patil College Of Engineering, Akurdi Pune.

Abstract:

This paper describes a novel multimodal image retrieval on mobile (smartphone) devices. It is designed for users who refining the recognized query by interactively composing image query using exemplary images, the user can easily find few natural multimodal interactions with his/her mobile devices. Our work focuses the mobile search experience and increases relevance of search results. It involves a natural interactive process through which user has to express their search intent very well.

Index Terms: Mobile visual search, Multimodal search, Mobile Devices (Smartphones), sketch based image retrieval (SBIR), Content Based Image Retrieval (CBIR)

1. Introduction

Image search is an emerging topic in both computer vision and information retrieval in many applications. The traditional desktop image search system with text queries have dominated the user behavior for long time. While now a days more and more consumers use mobile devices as their personal gadget for surfing on the Internet. Along this trend searching is becoming more pervasive and one of the most popular application on mobile devices. It is observed that search interest on mobile devices is more than that on desktop. Search initiated by mobile devices is always have strong purpose, which are largely related to local information such as local spots and local business like cinemas, restaurants, hospitals etc. Existing search technique is based on text based search and local map search. The user can type entity name or look up on online local map to find the target. However Image search is becoming pervasive as the development of content based and image retrieval. This enables user to capture photo using inbuilt camera and then initiate search queries about objects in visual proximity to the user. A small screen limits the presentation of searching results which requires top result is more relevant on the phone. However using text as search as a like TinEye [13] on PC query can hardly meet this end. The surroundings text of web images are not always accurate, user must know exact terms the annotator used in order to get required images as he wants. Actually, text annotations are language dependent and there are more images which have no text information on web repository. All these deficiencies can ruin user experience of text based image search on mobile devices. Comparing with the text search and map search visual multimodal search is still not very popular on the mobile phone. Main reason is that existing image search applications do not perfectly accommodate to the mobile and local oriented user intend, due to this searching is not very popular and user experience is hardly enjoyable. First of all typing is tedious job no matter tiny keyboard or touch screen are used. Even though voice queries are available on some mobile devices, there are still many cases that semantic and visual intent hardly be expressed by these description for search. Regarding content-based image search, one kind of famous products, including Google Image [12], TinEye [13] on PC and Google Goggles [15] on mobile phone can accept single images as search queries and return to the user similar images or even with information mined from their databases. With very large databases, these engines are able to achieve impressive results. However, to initiate such a visual search, the user must have an existed image on hand as a query. Moreover, it needs partially duplicate images or exact the same thing existing in the database.

ISSN 2348 – 7968

2. Literature survey

Recently, many systems developed that used multimodal query for image retrieval. Quickset was the first system that applied multimodal interaction with mobile systems [3], developed by the US Marine Corps. Speak4it local search application is another example [5], [6], where users generate mobile search queries by using multimodal commands that combines speech and drawing. Lee Chang proposed client-server architecture for mobile device to perform multimodal search. It mainly consists of four parts mobile Client, a carrier and forward server, a storage server, a media search server [7]. After that, many multimodal systems develop following client-server architecture. Xian Fan et al. proposed a system named Photo-to-Search to search information from the web on the go by using the captured photo [8]. When the user is attracted by any advertisement, they have captured image using camera and start searching process. To get such a query image is easier for camera phone users.

2.1. SBIR System

Sometimes when a user performs searching, example images will not be always at hand, which motivates sketch based image retrieval (S B I R) research that uses simpler Hand-drawn sketches as a query image. Among various query modalities, the sketch is the most challenging one. Compared with traditional search, sketch-based search is more accurate and convenient when a user's needed information is specific and complex, for example, if you want to find some picture of a beautiful pendant that you once saw in a shop. That query is usually too ambiguous to properly convey your search intention, but use sketch of pendant is simple. Users have to express their visual intent through sketches, but it's difficult for users without drawing ability. Yang Cao [11] proposed MindFinder system, which is the first interactive sketch-based multimodal image search engine. It enables 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.

2.2. CBIR System

Earlier researchers focus on searching for visually similar objects related to traditional Content Based Image Retrieval (CBIR) Technology [1]. On CBIR systems, image retrieval depends upon the content of images such as color, shape, and text information. However, CBIR-based approaches provide low precision because there is a big gap between high-level semantic concepts and low-level features. In recent year, I-SEARCH project [2] developed a multimodal search engine provides a novel unified framework for multimedia and multimodal content indexing, sharing, search and retrieval by using the concept of content objects (COs) [9].

2.3. Tap-to-search

A "tap-to-search" is other multimodal approach. It helps user to select only interested regions via "tap" actions on the mobile touch screen. Interested region candidates selected by using an Automatic segmentation technique [10].

3. Proposed System

Our work uses Android Application for information retrieval system. It deals with image retrieval using mobiles. The Objective is to design an efficient image aided visual search application in mobile devices on Android platform combine with local spot and scene search. Our work deals with image retrieval using mobiles. A connection is established between client and server. Then the client sends a captured image by entering correct IP address of the server and standard port number. Server accepts the client connection, reads the size of the image and image will download then the server checks the identical image, if the image is present it sends the information about the image to the client. The architecture is as shown in Fig1. And the modules used in the work are explained as below. It deals with two component server and client. A server is a process that is continuously running and waiting to be contacted by a client process. The server process is started on a computer system. In our work server is system, which accepts the client connection and start initializing response. Server contains the data base of our work, the flower images are taken as database.

ISSN 2348 – 7968

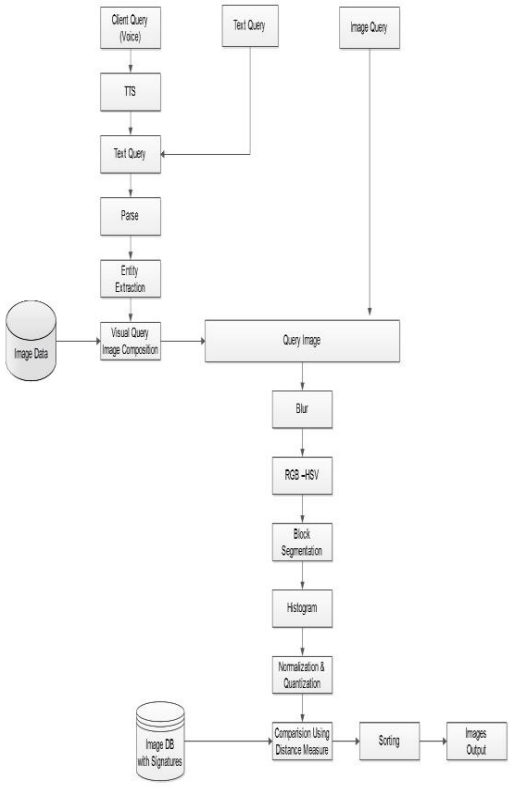

Fig 1. The Architecture of multimodal image retrieval system

3.1. Client-server Communication

The client-server model is a standard model for network applications. At a basic level, network-based systems consist of a server, client, and a media for communication as shown in Fig2. A computer running a program that makes a request for services is called client machine. A computer running a program that offers requested services from one or more clients is called server machine. The media for communication can be wired or wireless network.

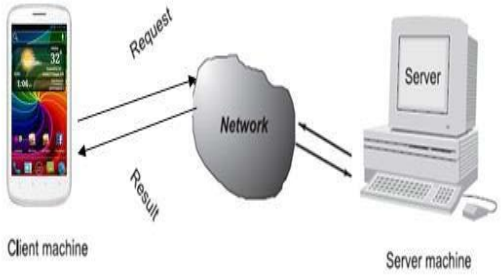

Fig 2. Client – Server communication

Mobile (client) details

- ☐ Model number LG-P715
- ☐ Android version 4.1.2(jelly bean)
- ☐ Camera 8MP
- ☐ Screen 10.9 cm (4.3 inch) LCD
- ☐ 1 GHz dual core
- ☐ Kernel version 3.40

3.2 Implementation View

3.2.1 Methodology

Several algorithms such as 1) Blurring an image algorithm 2) RGB TO HSV Model 3) RGB TO GRAYSCALE 4) Block Segmentation 5) Cosine Similarity 6) Porter Stemming algorithms are chosen for retrieving most relevant images.

A. ***Blurring an Image algorithm****:*

In image terms blurring means that each pixel in the source image gets spread over and mixed into surrounding pixels. Blurring an image reduces the sharpening effect, this makes the detection more accurate. There are two types of blurring an image: Gray Scale Blur and Color Blur

Steps:

- ☐ Traverse through entire input image array.
- ☐ Read individual pixel color value (24-bit).
- ☐ Split the color value into individual R, G and B 8-bit values.
- ☐ Calculate the RGB average of surrounding pixels and assign this average value to it.
- ☐ Repeat the above step for each pixel.

B. ***RGB To HSV Color Model***

HSV MODEL: H(hue) - Specify the position of pure color on wheel. S(Saturation)- Describe the how white the color is. E.g. pure red is fully saturated; tints of red have saturations less than 1. V (Value)-called as ‘lightness of color’. Describe intensity of color. Can be described as brightness in the color.

C. ***Block segmentation :***

Thresholding is the simplest method of image segmentation from a grayscale image, thresholding can be used to create binary images i.e. image with only black or white colors. It is usually used for feature extraction where required features of image are converted to white and everything else to black or vice-versa.

ISSN 2348 – 7968

D. ***Histogram Equalization***

Histogram equalization is a method of image processing of contrast adjustment using image's histogram.

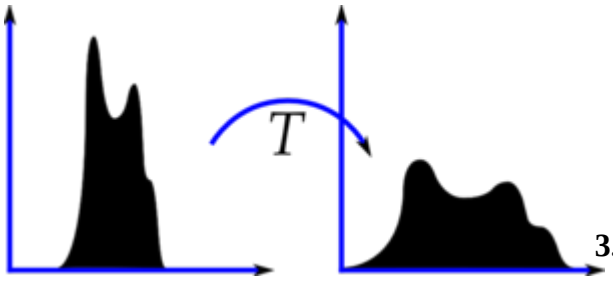

Fig 3.Histograms of image before and after equalization

This method usually increases global contrast of many images, especially when usable data of image is represented by close contrasted values. Through this adjustment intensities can be better distributed on the Histogram.

E. ***RGB To Grayscale Conversion***

Steps / Algorithm

- Traverse through entire input image array.
- Read individual pixel color value (24-bit).
- Split the color value into individual R, G and B 8-bit values.
- Calculate the grayscale component (8-bit) for given R, G and B pixels using a conversion formula.
- Compose a 24-bit pixel value from 8-bit grayscale value.
- Store the new value at same location in output image. Traverse Through Entire Image

Extract 8-bit R, G and B values from 24-bit Color Value
b = pix & 0xff;
g = (pix >> 8) & 0xff;
r = (pix >> 16) & 0xff;

3.2.2 Server Side Image Processing

On Server Side for each and every image store on server with proper tag value. Further tag value can be update or modify as per requirement.

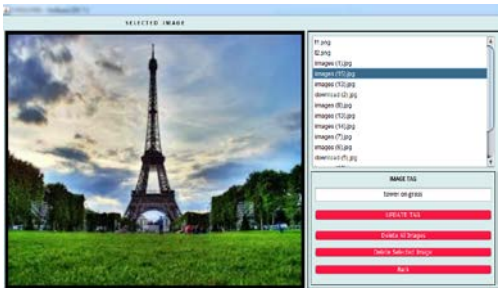

Fig 4.Managing Image on Server side

3.3.3 Client Side processing

On client side client send a request via three mode of searching viz. Text search, voice search and searching via image. In each type, request goes to server and search the given query via its tag value. To communicate with the server, client should know the server IP.

Fig 5. Searching Process

After gating getting the valid IP of server all the mode of searching proceed.

ISSN 2348 – 7968

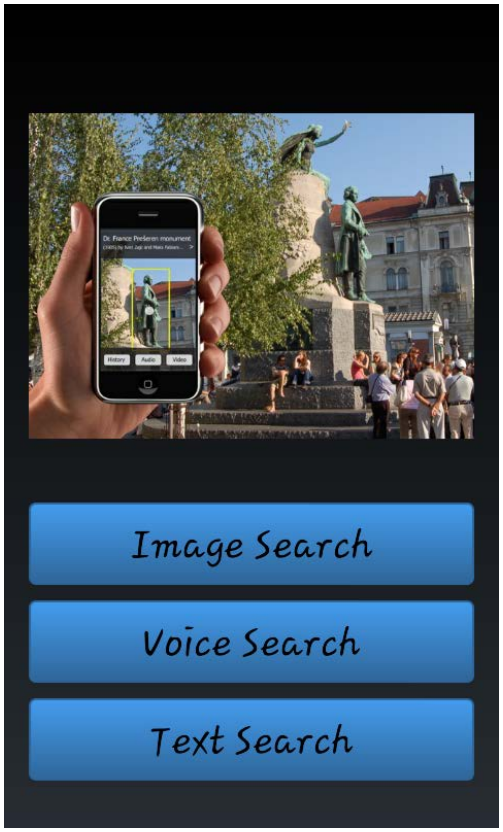

Fig 6. Mode of searching

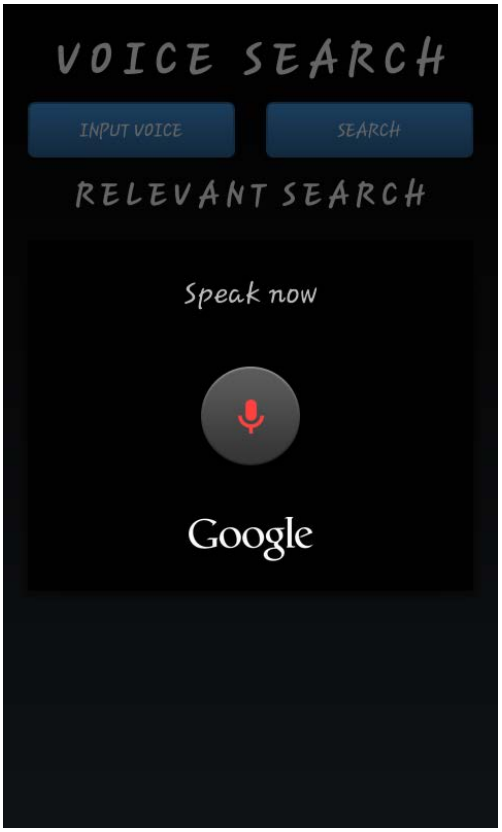

Fig 8.Voice Based Search

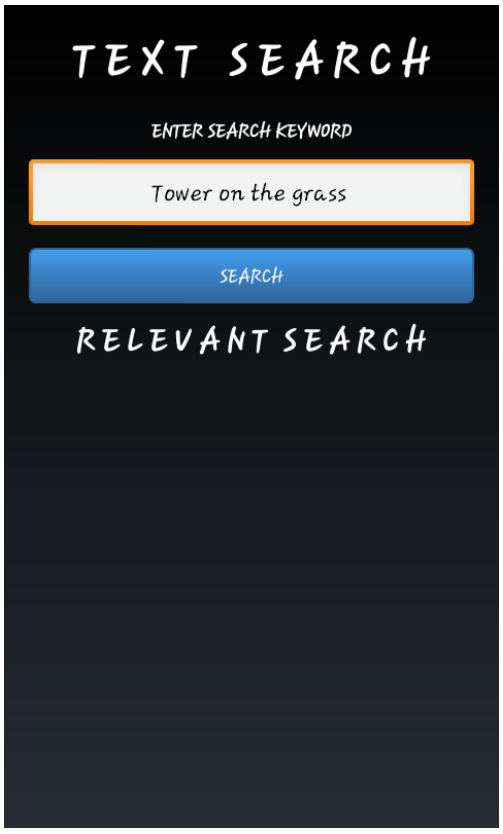

Fig 7.Text Based search

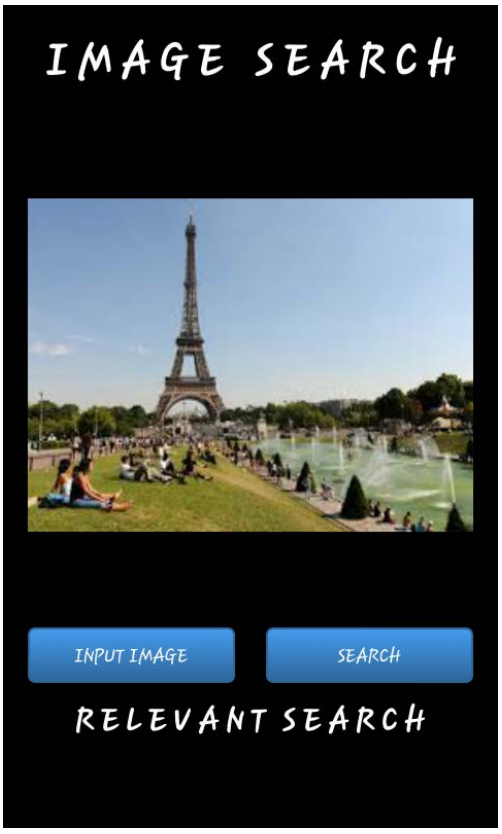

Fig 9.Image Based search

ISSN 2348 – 7968

4. Result and Discussion:

We can evaluate the effectiveness of the system forming methods, and compare the different applied methods, if we define metrics. Thus, we can determine which method works effectively in what circumstances, and when not.

Let in a test database containing N denotes pieces of images, P denotes length of retrieval list, from which Q pieces matter as relevant results and Z denotes the number of expected relevant hits. If we know this information, the following metrics can be calculated.

$$\text{precision} = \frac{\text{releventhits(Q)}}{\text{all hits(P)}} \quad (1)$$

Where the precision gives information about the relative effectiveness of the system.

$$\text{recall} = \frac{\text{releventhits(Q)}}{\text{expected hits (Z)}} \quad (2)$$

Where the recall gives information about the absolute accuracy of the system.

The number of all and expected hits is determined in each case of testing methods. The impact of multi-level retrieval to the efficiency of retrieval is measured, which conforms the importance of multi-level search. In addition, the ROC curves plot the true and false positive hit rate. The area under the curve reflects the efficiency of the method.

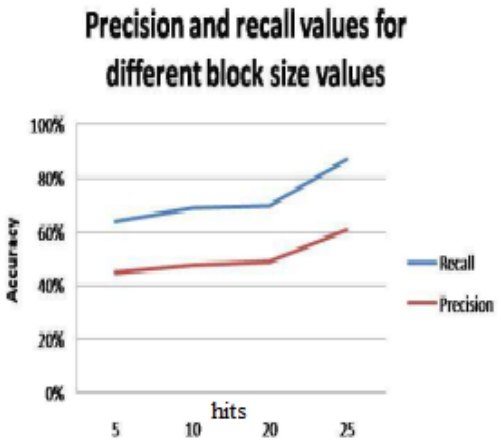

Fig 10. Effect of block size values

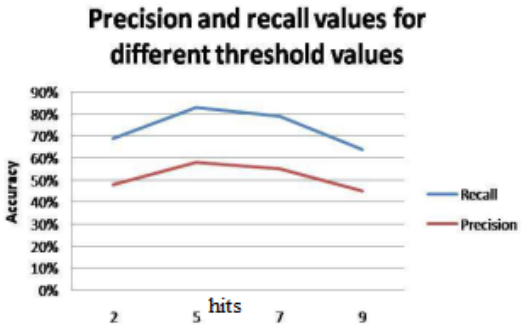

Fig 11. Effect of threshold values

5. Conclusions

In this paper, we have introduced image search system which allows user to search through mobile devices. It provides an image search scheme with composition of multiple exemplars. The visual query generated by the user can be effectively used to retrieve similar images by the proposed method. Compared to text-based retrieval system the performance of the proposed system is boosted. The user's search experience on mobile device is thus significantly improved by this image search system.

References

[1] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, and P. Yanker, "Query by image and video content: The QBIC system," IEEE Computer, vol. 28, no. 9, pp. 23-32, Sep. 1995

[2] Axenopoulos, A., Daras, P., Malassiotis, S., & Croce, V. (n.d.). ISEARCH : A Unified Framework for Multimodal Search and Retrieval, 130–141

[3] Philip R. Cohen, Michael Johnston, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen and Josh Clow. QuickSet: multimodal interaction for distributed applications. In MULTIMEDIA '97: Proceedings of the fifth ACM international conference on Multimedia, pages 31–40, New York, NY, USA, 1997. ACM.

[4] V. Chandrasekhar, D. M. Chen, A. Lin, G. Takacs, S. S. Tsai, N. M.Cheung, Y. Reznik, R.Grzeszczuk, and B.Girod, "Comparison of local feature descriptors for mobile visual

ISSN 2348 – 7968

search," in Proc. IEEE Int. Conf. Image Process., 2010, pp.

[5] Ehlen, P., Johnston, M., Ave, P., & Park, F. (2010). Speak4it : Multimodal Interaction for Local Search Categories and Subject Descriptors

[6] Michael Johnston and Patrick Ehlen, SPEAK4IT : Multimodal Interaction in the Wild, AT & T Labs Research, AT & T Labs. (2010), 147–148

[7] X. Xie, L. Lu, M. Jia, H. Li, F. Seide, and W.-Y. Ma. Mobile Search with Multimodal Queries. Proceedings of the IEEE, 96(4):589{601, April 2008

[8] Fan, X., Xie, X., Li, Z., Li, M., & Ma, W. (2005). Photo-to-Search : Using Multimodal Queries to Search the Web from Mobile Devices, 143–150

[9] Zahariadis T., Daras P., Bouwen J., Niebert N., Griffin D., Alvarez F., Camarillo G., "Towards a Content-Centric Internet", Towards the Future Internet - Emerging Trends from European Research, IOS Press, ISBN 978-1-60750- 539-6, pp. 227-236, Apr 2010

[10] Zhang, N., Mei, T., Hua, X.-S., Guan, L., & Li, S. (2011). Tap-tosearch: Interactive and contextual visual search on mobile devices.2011 IEEE 13th International Workshop on Multimedia Signal Processing,1–5.

[11] Wang, H., & Wang, C. (n.d.). MindFinder : Interactive Sketch-based Image Search, 1605–1608

[12] Google. [Online]. Available: http://www.google.com/

[13] TinEye. [Online]. Available: http://www.tineye.com/

[14] Siri. [Online]. Available: http://www.apple.com/iphone/features/ siri.html

[15] Google Goggles. [Online]. Available: http://www.google.com/mobile/ Goggles