Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 3159-0040
Vol. 2 Issue 9, September - 2015

Intelligent Visual Search For Mobile Devices

Komal V. Aher
ME Student, Department of Computer,
Sinhgad Institute of Technology,
Lonavala, India.
aher.komal@gmail.com

Sanjay B. Waykar
Department of Computer Engineering,
Sinhgad Institute of Technology,
Lonavala, India.
sbwaykar@gmail.com

***Abstract*—With recent technological growth, use of mobile devices is also increased, every one having smart phone with them and now days it became one of the necessary thing. Visual Search has become recent trend in fields like information retrieval and computer vision. So being a part of this trend, this paper gives an intelligent system for multimodal visual Search on smart phones which completely utilizing multitouch and multimodal utilities of smart phones. This visual search allows search using audio, text, image and composite images also. Proposed system is helpful for users who previously has snapshot in their brain or mind but don't have proper depiction to address it.**

The paper gives system which has facility to shape fused visual query to put across user's thoughts more obviously and to give more specific or significant outcome to end user. The anticipated algorithm will considerably obtain enhancement in many aspects. It also uses Context based picture retrieval method to get noteworthy outcomes. So system has ability to attain gain in terms of retrieval accuracy, performance and user satisfaction.)

Keywords*— *Mobile device, Visual search, Composite image, Histogram, color space, Quantization etc.

I. INTRODUCTION

If we turn around we will come to know as everyone is having cell phone with them and now mobile or smart phones became best choice for surfing on the internet. Moreover this fashion, searching is becoming pervasive and one of the most well-liked application on mobile devices [6]. User interfaces particularly for input types and search curiosity of users is completely different for the mobile devices and desktop computers. For mobile phone users existing search options incorporate text-based search and nearby or local map search. Moreover, photo-to-search is flattering continual as the spreading out in the computer vision and content-based image retrieval. That permits the user to take a snaps or pictures using the built-in camera on the mobile phone and then start search queries about things in visual immediacy to the user.

Still visual (image and video) search is not too much accepted on the mobile devices as arbitrator against with text search, map search, and photo-to-search. The major concern is why such image search applications are not accepted on mobile device because small screen of device and the existing search applications do not absolutely accommodate to the mobile and local oriented user intention. So in such conditions such irrelevant images spoil the results and ruin the user experience, visual-aided tools can mostly improve the relevance of search results and the user experience [3]. Let's think about such a situation in which the user has no idea of the name of target but can only depict its particular appearance, that only with a sight or general portrait in the user's mind, such kinds of searches are not easy under current text-based search condition. But with the assist of visual aids, search for images not only based on text but also based on image content, these tasks can be more significant.

Visual Search has become recent trend in fields like information retrieval and computer vision. In the proposed work we facilitate an intelligent system for multimodal visual Search on smart phones which completely utilizing multitouch and multimodal utilities of smart phones. If the users have an image at their side, they can use it directly as a query and find matching images in dataset. Otherwise users can easily create merged images as their search query by naturally interacting with the phone through voice and multi-touch. So the system permits users to convey their implicit and explicit search intent well. We have designed a multimodal visual search system to carry out different types of queries from mobile phones and expressing user's information needs in a better way. For improved search contextual information also added to the system. As a result, a powerful visual search system with visual support is designed. The proposed system gives compliment to existing systems and will give user friendly inter face and satisfactory results to user conceptual query.

II. RELATED WORK

A. Visual Search

Visual search is a kind of perceptual task requiring consideration that typically includes an active scan of the visual environment for a particular object or the target among other objects or the distracters. There are some presented systems as:

1) Text to Search:

Google and Bing are traditional text-based search engines are available on mobile devices. But, widespread text queries are neither user-friendly on phone, nor machine-friendly for search engine. The truth is that the mobile users use only 2.6 terms on

average for search [6], which can hardly express their search intent.

2) Voice to Search:

As we know typing is a tiresome task on the phone no matter whether a tiny keyboard or a touch screen is used. Although speech inputs are available on some devices, still many cases that semantic and visual purpose cannot plainly articulated by these descriptions for search [3]. However, the users typically have to accept some ideal images amidst much other inappropriate result. In such cases where irrelevant images spoil the results as well as user experience, visual-aided tools can largely boost up the relevance of search results and the user experience.

As the speech recognition became mature, Smartphone applications using speech recognition rapidly grows. The most popular application is Apple Siri [11], which combines speech recognition, natural language understanding and knowledge based searching techniques, with an ability to make a speech conversation with the phone and get information and knowledge from it. The user can ask the phone for anything by only speech and get multimedia answers.

3) Sketch to search:

In Sketch-based image search hand-drawn sketches were used to search for satisfied images. But with these systems it is very tough to convey user intention and is not easy for users without drawing practice to use such application. [10] [13]

In [7] the authors build a Sketch2Photo system that uses simple text-annotated line sketch to automatically synthesize realistic images. However, their work focuses on image composing instead of visual search.

4) Photo to Search:

It permits the user to capture photos using the in-built camera of the mobile phone and then start search queries regarding objects in visual proximity to the user. This advance offers various applications such as comparison shopping, identifying products, finding information of movies, buildings, compact CDs, print media, real estate, artworks, etc. [3]

Photo-to-search applications are become imperative on mobile phones. These techniques enable users to seek for what they see by taking a photo on the go. As Google Goggles, Point and Find [8], and Snaptell [9] are good examples. These applications seem for the precise partial duplicate images in their records and provide the users with associated information of the query images. But, the search is only accessible for some vertical domains, such as CD covers, products, landmarks etc., where the partial duplicate images of the query image have been indexed in their records.

B. Concept Based Image Retrieval

It is a conventional approach to any form of image search, as concept-based image indexing. Also called as description-based or text-based image retrieval, this kind of search refers to keywords, tags, captions, matter headings or natural language text for the indexing and retrieval of text-based images. For years now, SEOs and digital marketers have been optimizing images so that explore engines like Google could recognize and appropriately index visual content [14].

C. Content Based Image Retrieval(CBIR)

With CBIR, search engines examine the visual content of the image (pixels) rather than the metadata. In this the scheme of "content" may refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is gaining popularity because of the inefficiencies and boundaries inherent with metadata-based image retrieval. Optimizing for text-based retrieval can be time overwhelming and create inadvertent ambiguities. However, until recently, many image retrieval systems, such as Google-image search, were exclusively text based [14].

D. Reverse Image Search

Reverse image search is a CBIR query system that involves providing the search engine with a model image to base its query on. Reverse image search allows users to determine content that are related to a specific model image, popularity of an image, and find out manipulated versions and derivative works. Different implementations of CBIR formulate use of different types of user queries. Examples are Google Image search and Tin Eye [14] [16].

III. PROPOSED SYSTEM

A novel system is designed to deal with troubles faced by user while dealing with mobile device to do visual search. Here one mobile based application is designed for visual searching with user interactive and game like interface to provide more appropriate outcomes to user. The proposed system is helpful to users who can't convey their search intend clearly.

In a proposed work, standard structural design is followed as client-server. Mobile device with android OS is treated as client and computer or laptop as Server. Wherein visual database is preserved and maintained on Server by administrator. At Client side, there are four modes are given to initiate image search which is a combination of all existing systems. Proposed system general idea is shown below which will simply do visual search task with user friendly manner:

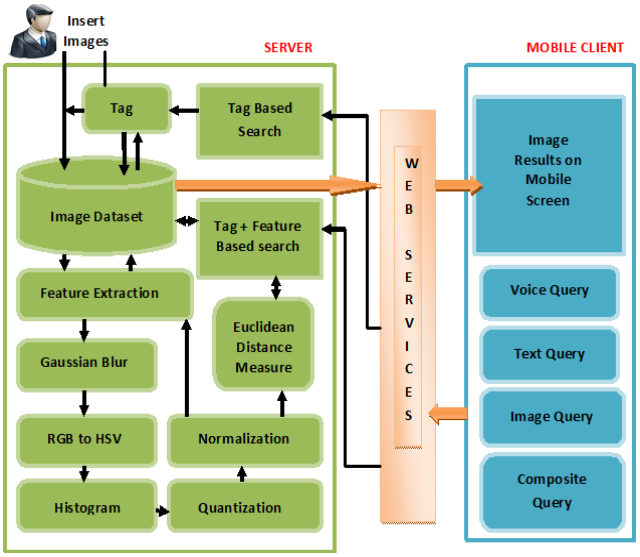

Fig. proposed System Architecture

A. Server Side Implementation:

The architecture mainly describes, at server side image dataset is created and maintained, wherein images are stored as per sorting by Admin. There are two main things are considered while storing images into database- Image label or tag and its Features. So when any image is inserted into dataset at first tag is assigned to it and the features are extracted for each and every image.
Features of Images stored into the dataset are extracted from images. For that each image is passed through various image processing algorithms. Various operations are performed on each and every pixel of image such as Blurring, RGB to HSV, Color histogram calculation, Quantization and then Finally Normalization. All these algorithms are given as:

a. Image Blurring:

In Image Blurring it reduces the sharpened effects of image; which aid to make the recognition more accurate. In blurring every pixel in the source image gets stretch over and mixed into adjacent pixels. Different steps for image blurring are:

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

Gaussian Blur is one of the most effective methods.

b. RGB to HSV Color Model:

After image blur pixels of image are converted from RGB to HSV model to make it easier to perform any operations. The HSV stands for the Hue, Saturation and Value, gives the perception illustration as per human visual aspect. Hue specifies the color type its range as 0 to 360. Saturation gives the vibrancy of the color ranges from 0 to 100%, and it also called as purity. Value, the brightness of the color ranges from 0 to 100%. RGB to HSV conversion steps are as follows:

1. Find minimum value of basic R,G,B
   Min=min(R, G, B)
2. Find maximum value of R, G, B
   Max=max(R, G, B)
3. Calculate HSV from RGB
   temp=Max-Min
   Value=Max
   If (Value==0) then
   Hue = Sat = 0
   Else
   Sat=255 * (temp)/Value
   If (Sat==0) then
   Hue=0
   Else if (Max==R) then
   Hue= 0 + 43 * (G – B)/temp
   Else if (Max==G) then
   Hue= 85 + 43 * (R – B)/temp
   Else if (Max==B) then
   Hue= 171 + 43 * (R – G)/temp
   If (H < 0)
   H = H + 255
   Set Pixel.

It is observed that HSV model accuracy is higher as compared to RGB model. [12]

c. Histogram:

Color histogram represents the distribution of the composition of colors in the image. The histogram consists of bins where each bin defines a small range of pixel values. The value stored in each bin is the number of pixels in the image that are within the range. These ranges represent different level of intensity for each color component. The values in each bin are normalized by dividing with the total number of pixels in the image.
Then, by counting the number of pixels in each of the bins, we get the color histogram of the image.

d. Quantization:

The quantization of the number of colors into several bins is done in order to decrease the number of colors used in image retrieval. Quantization reduces the number of colors used in an image. While computing the pixels of diverse colors in an image, if the color space is outsized, then first segregate the color space into certain numbers of small intervals. Each of the intervals is called a bin. This process is called color quantization.

e. Normalization:

Normalization is a process that changes the range of pixel intensity values. It sometimes called contrast stretching or histogram stretching. The purpose of normalization in the various applications is usually to bring the image, or other type of signal, into a range that is more familiar or normal to the senses.

f. Distance Measure Estimation:

Similarity measure is real value function that quantifies similarity between two objects, so here for similarity calculation between images Euclidean distance measure is used.

B. Client Side Implementation:

At another side on mobile device user can input his query in various manner. To communicate with the server, client should know the server IP. Basically there are four ways are provided to initiate image search:

a. Text input:

In a direct text input, user can give query as single keyword or multiple keywords. According to text input tag based images are retrieved and if resultant images are not specific to query then current results are forwarded as composite query of images and then tag and feature based image retrieval initiated then it results into more relevant images to the queries.

Image result for Voice And Text Query:

- Input: Text Query
- Output: Result Images
- Algorithm:
  1: Give text query as input
  2: Tag based image search from database as query entity.
  3: arrange resultant images on mobile screen
  If (user is satisfied) then,
  Done
  Else
  Make composite query using multiple tags
  4: According to tags image collage is shown on mobile screen
  5: Image collage is given as query
  6: Feature based image retrieval at server
  7: Most similar results are shown on mobile screen
  8: Relevant image result.

b. Voice input:

To convert speech given by user as input is converted into text by using standard Google's speech recognition tool and result in more accurate text. As every android mobile have inbuilt speech to text conversion tool so proposed work uses the same.

Image result for Voice And Text Query:

- Input: Natural voice or text Query
- Output: Exemplar Images
- Algorithm:
  1: give text or voice query as input
  2: Tag based image search from database as query entity.
  3: arrange resultant images on mobile screen
  4: Output images as an exemplars.

As voice input is converted into text, that text is parsed and treated as tag. And according to tags image results are retrieved form predefined image dataset. And results as exemplars are shown to user.

c. Image input:

Image result for Voice And Text Query

- Input: single image
- Output: Similar images
- Algorithm:
  1: give image as input
  2: Feature Extraction by blurring, color conversion, histogram etc.
  3: Tag + feature based image retrieval started
  4: Similar results on screen.

d. Merged Images input:

In this user can select two image from results or pre stored image from device. Then that image is passed through image processing algorithms for feature extraction then feature based similarity is checked with help of Euclidean distance measure and finally relevant results are shown to user.

Composite Query Processing:

- Input: result images, photo
- Output: Final resultant images
- Algorithm:
  1: Check for input image or predefined photo
  or
  Composite visual query
  2: Segmentation Based Image representation using Histogram
  3: Color feature extraction
  4: Finding most similar images from dataset based on feature
  5: Apply Euclidean distance measure
  6: Rank or sort results
  7: Final relevant output Images

.

IV. Results And Discussion

For proposed image search application, results are calculated based on two parameters as time and accuracy. Time estimation for each type of search for proposed system is as follows:

Process	Time (ms)
Voice Search Activity	7168
Text Search Activity	1205
Image Input Result	5337
Image search Activity	1403
Composite Image search activity	1582

Table 1.Time Estimation

The accuracy is measured using precision and recall parameters. Some testing results are shown into below table:

Type	Actual Objects	Retrieved Objects	Correct Retrieved Objects
Image Search	11	8	5
Voice Search	11	2	2
Text+Image Search	11	7	6

Table 2: Searching Results

Using above values precision and recall is calculated as follows:

Precision = (Relevant Intersect Retrieved) / Retrieved
= Correct Retrieved Object / Retrieved Objects

Recall = (Relevant Intersect Retrieved) / Relevant
= Correct Retrieved Object / Actual Objects

Type	**Precision**	**Recall**
Voice	0.625	0.454545455
Text + Image	1	0.181818182
Image	0.857142857	0.545454545
Total	**0.827380952**	**0.393939394**
Accuracy (%)	**0.3939393934**	

Table 3: Precision and Recall Calculation

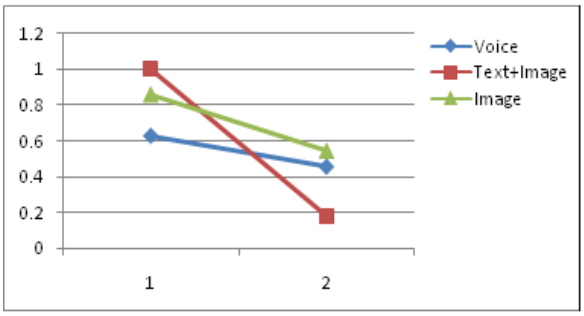

Fig. precision and recall for different inputs

The proposed system is working efficiently on mobile phones and accommodate mobile screen with completely utilizing multitouch and multimodal functionalities of device. As per above outcome user will find satisfied by image results generated by proposed system.

V. CONCLUSION

The paper proposes a system which completely utilizing multimodal and multitouch functionalities of mobile devices and gives game like interface for image search. Thus, the paper present system which has ability to form merged or composite visual query to convey user intent more clearly which helps to give more precise or relevant outcome to user. The proposed algorithm will significantly improve in different aspects. System also use context based Image retrieval schema to give relevant results. So system is able to achieve gain in terms of search performance, accuracy and satisfactory results to their imaginary queries.

ACKNOWLEDGMENT

I would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. I am thankful to the authorities of Savitribai Phule University of Pune. I'm also thankful to reviewer for their valuable suggestions.

REFERENCES

[1] K. F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "An efficient and effective region-based image retrieval framework," IEEE Trans. Image Process., vol. 13, no. 5, pp. 699–709, 2004.

[2] B. Girod, V. Chandrasekhar, D. Chen, N. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. Tsai, and R. Vedantham, " Mobile visual search," IEEE Signal Process. Mag., vol. 28, no. 4, pp. 61 76, 2011.

[3] Houqiang Li, Yang Wang, Tao Mei, Jingdong Wang,and Shipeng Li,"Interactive Multimodal Visual Search on Mobile Device", IEEE Transactions on Multimedia, VOL. 15, NO. 3, April 2013.

[4] C.Wang, Z. Li, and L. Zhang, " MindFinder: Image search by interactive sketching and tagging," in Proc. Int. Conf. World Wide Web,pp. 1309 1312, 2010.

[5] Y.Wang, T. Mei, J. Wang,H. Li, and S. Li, "JIGSAW: Interactive mobile visual search with multimodal queries," Proc. ACM Multimedia, pp. 73 82, 2011.

[6] K. Church, B. Smyth, P. Cotter, and K. Bradley, "Mobile information access: A study of emerging search behavior on the mobile internet," ACM Trans. Web, vol. 1, no. 1, May 2007.

[7] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet image montage," *ACM* Trans. Graph., vol. 28, no. 5, pp. 124:1–124:10, Dec. 2009.

[8] NOKIA Point and Find. Link: http://pointandfind.nokia.com/

[9] Link: http://www.snaptell.com/

[10] Y. Cao, H. Wang, C. Wang, Z. Li, L. Zhang, and L. Zhang," MindFinder: Interactive sketch-based image search on millions of images", in Proc. ACM Multimedia, pp. 1605 1608, 2010.

[11] Link: http://www.apple.com/iphone/features/siri.html

[12] Simardeep Kaur and Dr. Vijay Kumar Banga, "Content Based Image Retrieval: Survey and Comparison between RGB and HSV model," International Journal of Engineering Trends and Technology (IJETT) - Volume4Issue4- April 2013.

[13] M. Eitz,K.Hildebrand, T. Boubekeur, and M. Alexa, "An evaluation of descriptors for large-scale image retrieval from sketched feature lines,"

Comput. Graphics, vol. 34, no. 5, pp. 482–498, 2010.
[14] Link : http://www.google.com/
[15] Y. Cao, H. Wang, C. Wang, Z. Li, and L. Zhang, “MindFinder: interactive sketch-based image search on millions of images”, in Proc.ACM Multimedia, 2010.
[16] Link: http://www.tineye.com/
[17] H. Xu, J. Wang, X. Hua, and S. Li, " Image search by concept map," in Proc. ACM SIGIR, 2010, pp. 275 282.
[18] 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, pp. 227 236, Apr 2010.