Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 12, December 2014, pg.83 – 89

RESEARCH ARTICLE

Interactive Robust Multitudinous Visual Search on Mobile Devices

Ms. Roshani E. Bagul, Prof. K.P.Gaikwad
Department of Computer Engg, S.N.D.C.O.E.R.C
Yeola, Dist-Nasik, Savitribai Phule University of Pune, Pune, Maharashtra, India
roshanibgl1@gmail.com

Abstract— This paper describes a novel interactive multitudinous visual search system for multiple type of input e.g. text, image, voice or any combination of that inputs using Geotag on mobile devices. The system, jointly search for Image, Speech, And text, takes full benefits of the multiple type of input and get best user interactions on mobile devices. It is designed for users who have number of imaginations in their mind but unable to describe their expectations and cannot get desired results for entered query. The proposed system is able to achieve best search performance and is ten times faster. We are developing a system to search result using Geotag by which tracing location which will show best desired result (location based search), with all possible similar images. This is the main advantage of proposed system. We are adding large vocabulary (trained dataset) for getting maximum interesting result for multitudinous input.

Keywords- Interactive search, multitudinous, location based search

I. INTRODUCTION

Problem Statement

Now a days searching is becoming pervasive and is one of the most popular applications on mobile phones. People are more and more addicted to search on their phones. It is reported that one-third of search queries will come from smart phones by 2014. And it is tedious task of typing a query in text format through desktop devices.

To facilitate visual search on mobile devices, our work focuses about more natural way to formulate a visual query, taking full advantage of multitudinous and multi-touch interactions on mobile devices. In Existing system, firstly user takes a picture and then performs similar image searches in various vertical domains (i.e., capture-to-search mode). However, in many cases, the user's search intent is implicit and cannot be represented through capturing the surroundings. The user, nevertheless, can express his/her intent via a piece of voice description.

Motivation of the Proposed Approach

We introduce a new multimodal search system for image retrieval that helpful in two different situations. Consider an example a user move to an unfamiliar place and takes food from one unknown restaurant. To visit the same restaurant next time. 1) He simply takes a picture of that one.2) another situation is that he forgets to take pictures of the restaurant. After returning only he remembered. He also has no idea of the name of the restaurant, but can describe its particular appearance such as "a Chinese restaurant with white door, two red lamps lions, and many red pillars in front. The proposed system handles these two situations. In first case system uses an image query such as a captured photo of the restaurant and start searching process and retrieve similar images from the web. In the second case, user's doesn't have an existing image, but the user can generate an image query by giving speech input to the system, that represent picture described in the user's mind. Earlier developed 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.

II. LITERATURE SURVEY

There are number of system deployed on a real-world phone system, evaluated the performance on one million images, As a mobile visual search system, the most related works include many multimedia search apps available on mobile devices. Directly applying text-based search to mobile visual search is straightforward.

Traditional text-based search-
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 the mobile users use only 2.6 terms on average for search [8], which can hardly express their search intent. As the pervasive mobile phones support 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 developments in mobile phone applications for multimedia search.

Apps	**Features**
Goggles	Product, cover, landmark, name card
Digimarc Discover	Print, article, ads
Point and Find	Place, 2D barcode
Snapnow	MMS, email, print, broadcast
Kooaba	Media cover, print

Table 1 Visual search applications providing number of features

Traditional Speech based search-
As the speech recognition became mature, phone applications using speech recognition rapidly grows recently. The most representative application is Apple Siri [5], 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.

Photo-to-search-
Photo-to-search applications also became pervasive on mobile phones. Such applications enable users to search for what they see by taking a photo on the go. As we have summarized in Table I, Google Goggles, Point and Find [4], and Snaptell [6] are good examples in this field. These applications search for the exact partial duplicate images in their database and provide the users with related information of the query images. However, the search is only available for some vertical domains, such as products, landmarks, CD covers, and etc., where the partial duplicate images of the query image have been indexed in their database.
In academic circles, there is not a big difference from industry. Researchers for mobile search also focus mainly on photo-to-search techniques. The research topics include visual feature design, database indexing, and transmission.

Content based search-
Regarding content-based image search, one kind of famous products, including Google Image, TinEye on PC, and Google Goggles 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. Another kind of image search engines designed for desktop, including GazoPa and some other sketch-based image search researches, use hand-drawn sketches to search for satisfied images. Though sketch-based search allows users to express their visual intent in some way, it can hardly develop complex meanings and is difficult to use for users without drawing experience. MindFinder and Concept Map also provide visual aids to search for images. In these works, visually and semantically similar images are retrieved by multiple exemplary image patches

Feature Descriptors-

Traditional features such as MPEG-7 image signature, SIFT,Speeded Up Robust Feature (SURF) [1], and Oriented FAST and Rotated BRIEF (ORB) [3] are widely used in such visual search systems because of their invariance to illumination, scale and rotation. Moreover, compressed visual descriptors are proposed to accommodate the limited processing speed and narrow bandwidth on the phone.

Chandrasekhar *et al.*discussed their compression as well as proposed a new feature of Compressed Histogram of Gradients (CHoG) [7]. It can quantize and encode gradient histogram with Huffman and Gagie trees to produce very low bit-rate descriptors. Besides, various systems with optimized technique and better indexing are developed to search for landmarks [8], books [9], CD covers [11] etc. In [10], indexing with bundle of features is introduced in a poster and CD cover retrieval system for mobile visual search. In [11], different local descriptors are compared in a system of CD cover search. SIFT is widely accepted as the best performed feature and CHoG has an advantage in low-bit transmission. Bernd Girod *et al.* gave a comprehensive overview of photo-to-search in [12], from the architecture of an efficient mobile system to the framework of an image recognition algorithm. Other techniques are also used in visual search such as barcodes and OCR [9], [13].

III.EXISTING SYSTEM

Currently there are many services offering for retrieval of images using text query only. Limitation of this system is the output of such systems is in bulk mode without the exact accuracy. Existing system does not take input query in the form of audio and combination of audio and image, or any combination of available inputs for searching process. So, most of time users will get the result which was not expected by them. Main issue of present theory, it can not combine low level features into mid-level features and due to lack of large vocabulary size can not able to give more search results as per user expectations. And it can not show more similar search results for entered query.

The existing system, joint search with Image, Speech, And Word Plus (JIGSAW+) takes input in different format but not in combination of them. And do not have facility of location based search. Existing text-based visual search system still not sufficiently improved over the years to accommodate the new emerging demand of mobile users. Searching on one's phone is becoming manifested.

IV.PROPOSED SYSTEM

It is designed for users who already have pictures in their minds but have no precise descriptions or names to address them. By describing it using speech and then refining the recognized query by interactively composing a visual query using exemplary images, the user can easily find the desired images through a few natural multimodal interactions with his/her mobile device. Whenever user will give search query in terms of audio then we convert it into text using ASR (Automatic Sound Recognition), result in text is passed to Image Search engine using JSON. And whenever user will give search query in terms of image including text then we extract the text from image using TESSER and OCR(Object Character Recognition) algorithm for fetching appropriate results. We are also implementing a module which gives us exact image retrieval on the basis of its Geotag with its input image. In android device, it is possible to add the Geotag of the current location. We are using that Geotag for further image retrieval system.

As a next step, our proposed system add a geotag to image for efficient search process, geotagging also may need into augmented reality searches. In one example, images that are geotagged may show up in visual searches of streets and landmarks using a smartphone camera. These tags use latitude and longitude information, or regions, in order to help nail down the location of the content. The geotagged images are going to become much more relevant for presenting information. Using the restaurant example, imagine that someone with a smartphone does a location-based search for a restaurant in particular city. Then the search result gives all possible related images of restaurant of that city using geotag.

Proposed work

The proposed system handles three situations -

In first case system uses an image query such as a captured photo of the railway and start searching process and retrieve similar images from the web. In the second case, user’s doesn’t have an existing image, but the user can generate an image query by giving speech input to the system, that represent picture described in the user's mind. And in third case system fetch geotagged images to search intended images for particular area. Our system uses speech recognition engine service to convert user speech to text input. Then keywords extracted from the text. Based on these keywords, users can start searching process. The images correspond to each keyword, then arranged on canvas and generate a composite image query that used as searching query.

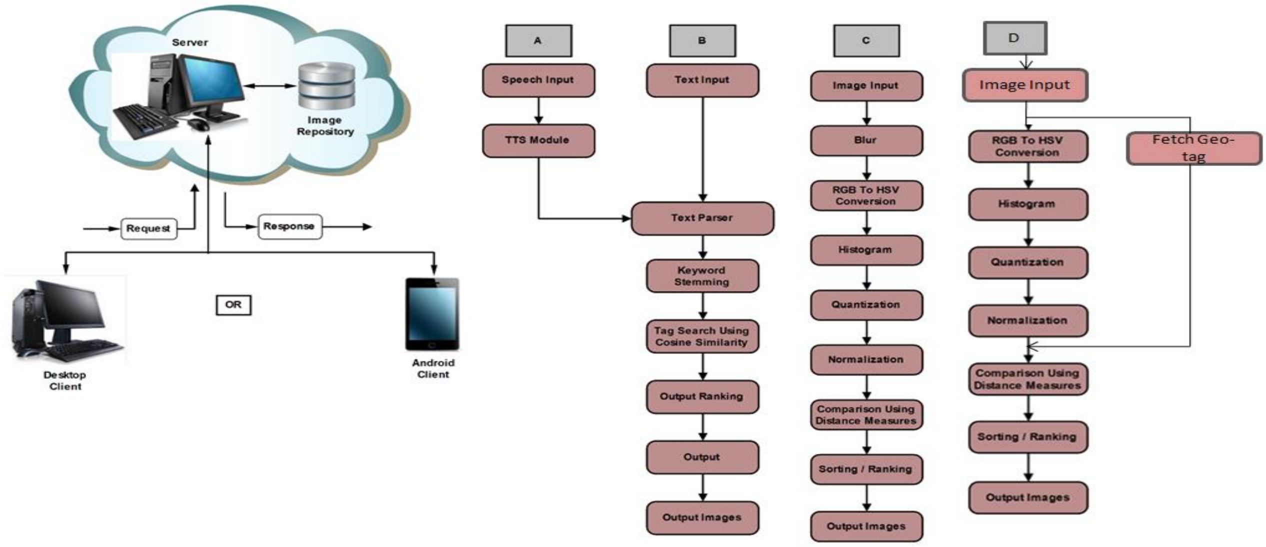

Fig. 01 system architecture

V.SYSTEM DESIGN

A.Visual Query generation

Different from traditional text-based image search system, in the proposed image search system, the user can use a visual query to further articulate the user's visual intent.

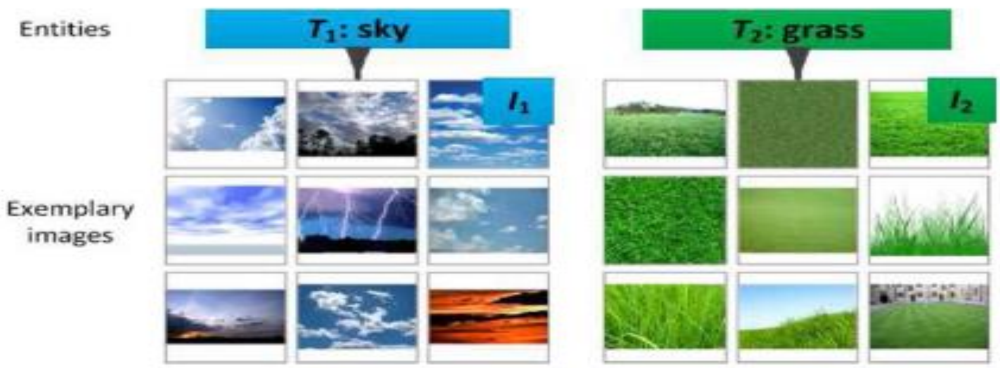

Fig 02 An example of processing a visual query with multiple exemplars from an initial query with multiple entities . Each entity corresponds to one exemplary images in the composite visual query. There are two recognized entities: "sky," and "grass". For each entity, there is a list of candidate exemplary images. The user can select any image in each list. The selected exemplary images are denoted as I1 and I2.

a.Visual Exemplar Generation

A visual exemplar database should be ready to provide exemplary images to the user according to their keywords. As a result, plenty of exemplary images are generated offline. Given a keyword, a set of images are first downloaded from the Internet by inquiring a image search engine. Usually, top results from a Internet search engine for simple concepts are reliable than many other resources (e.g., Flickr).

b.Composite Visual Query Generation

For each keyword, the system offers to the user with several exemplary images according to the clustering results. The user can choose one image under each keyword and put it onto the canvas region on the phone screen. The user can both reposition and resize the exemplary images until satisfied. Fig. 4 gives some examples of the composite visual query with different number of exemplary images. Finally, by choosing and positioning single or multiple exemplary images, a composite visual query is generated by the user.

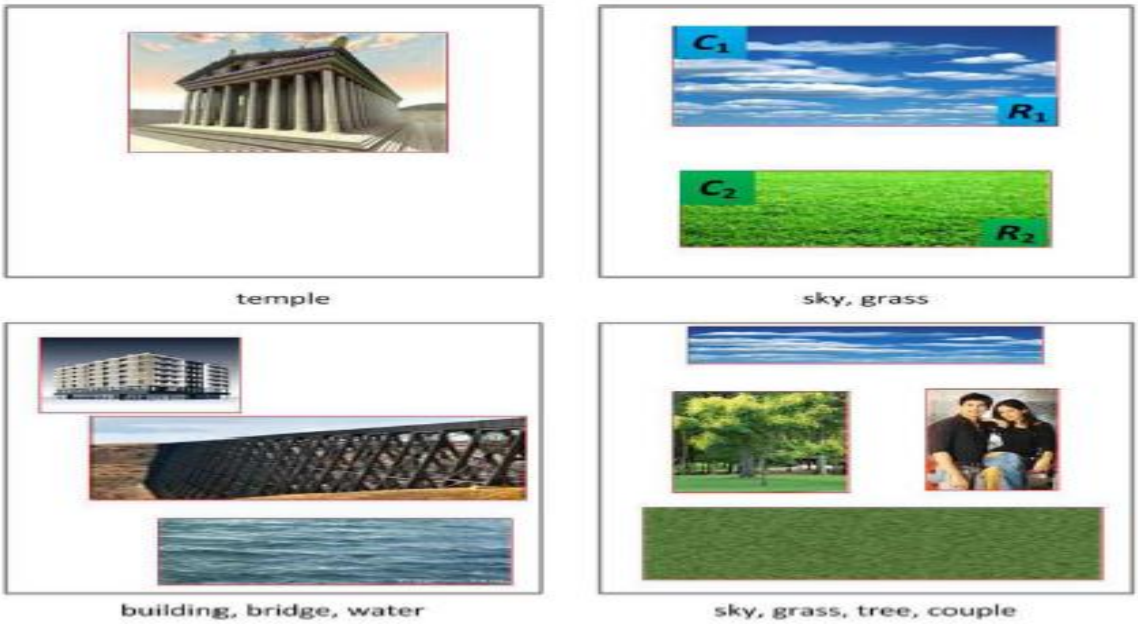

Fig 03 . Some examples of visual queries. Take a second visual query for example, composite visual query is made up by the selected image I1 and I2, where the position and the size of each exemplary image on the canvas are adjusted by the user.

C.Visual Representation of Images

As a content-based image retrieval system, a series of features should be first extracted from images, where color features and texture features are widely used.

1) Color Feature Extraction: Before color feature extraction, the images are over-segmented using algorithm by Felzenszwald and Huttenlocher described in [13]. This algorithm has been widely used in image content analysis such as [21] and [3]. This algorithm is easy to use and has advantage in speed, though other methods may bring about better results.

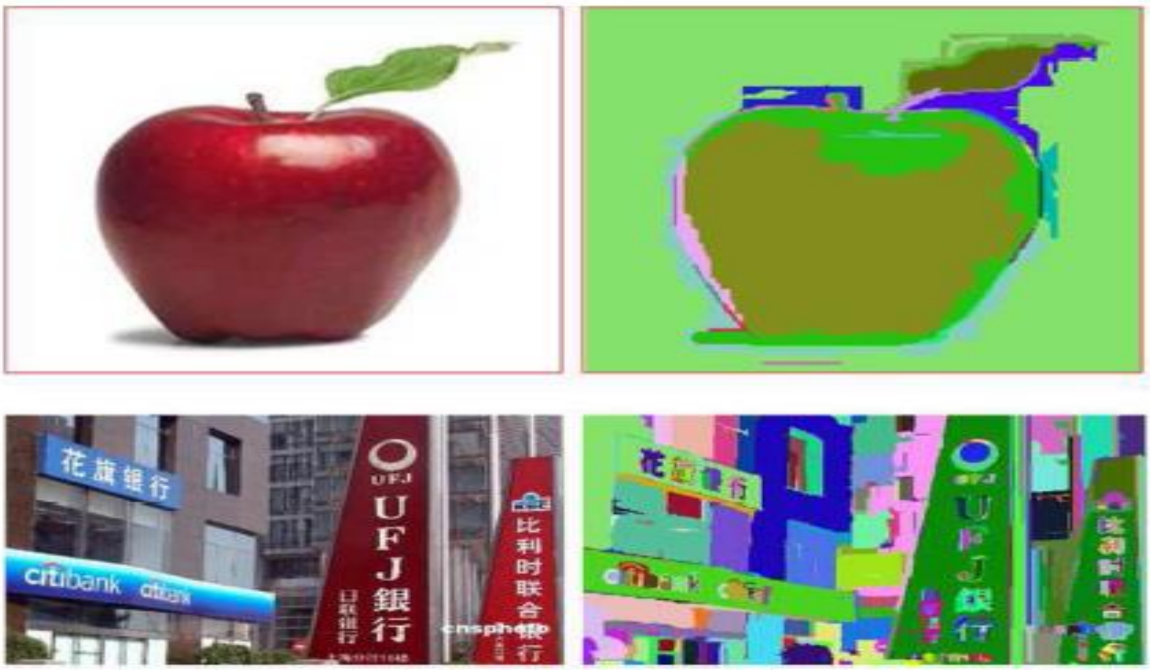

Fig 04 .Examples of the over segmentation.

2) Texture Feature Extraction*:* Beside the homogeneous regions, we also extract some unique texture regions as secondary feature. We use the most widely used local feature of SIFT to capture local texture patterns in the image. To reduce computing and noise, only prominent regions are used instead of homogeneous regions used in color feature extraction, or grid-based dense sampling. Prominent regions are detected using Difference of Gaussian (DoG) detector. By doing this, the size of each region can be also optimized via scale selection.

3) Descriptor Quantization*:* Bag-of-word model is used for both color and SIFT features to handle the uncertainty of the number of descriptors in the image. Codebooks are generated separately for color and SIFT feature, so that each descriptor is quantized to a visual code.

D. Multiple-Exemplar Relevance

After the composite visual query is submitted by the user, a search process starts at the server end. Images are favored which contain all the exemplars as well as spatially consist with the exemplars' geometry structure. As a result, the search procedure can be decomposed into three stages: 1) searching images in the database for each exemplar, 2) merge the results from different exemplars' results, and 3) check the spatial consistency of all the exemplars in the result images.

E. Indexing and retrieval

Million-scale images are indexed by the extracted visual words in a reverted index. Each image is represented as a series of visual codes first, and then a table mapping visual codes to images is constructed. The same visual codes are followed by lists of entries. Each entry keeps the information of a unique piece of an image including image ID, center point location and weight. Since the amount of visual similar but semantic dissimilar images is overwhelming and hard to control, the keywords are used first to mark all images that are relevance to any keyword by text-based retrieval. By fetching feature, images are ranked, and represented. This gives most intended result to user.

*Advantages***:**

1. Compared to text-based retrieval system the performance of the proposed system is boosted.
2. The user's search experience on mobile device is thus significantly improved by this game-like image search system.
3. Due to Geotag, region based search will be possible, and user will get related most relevant results.

VI. CONCLUSIONS

The proposed interactive multimodal image search system on mobile phone will helps users to express their needed information implicitly and explicitly. Only the user has known what information they are looking for, so expressing their information need play very important role in the search process. Proposed system gives a better way to express their visual intent than other existing systems, it provides more relevant search results, especially in case where users can have a partial picture description in mind. Thus, the user's search experience on mobile device is significantly improved by this image search system.

REFERENCES

[1] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded-up robust features," in *Proc. ECCV*, 2008, vol. 110, no. 3, pp. 346–359.

[2] 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.

[3] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An effi-cient alternative to SIFT or SURF," in *Proc. IEEE Int. Conf. Comput.Vis.*, Nov. 2011, pp. 2564–2571.

[4] NOKIA Point and Find. [Online]. Available: http://pointandfind.nokia.com/

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

[6] SnapTell. [Online]. Available: http://www.snaptell.com/

[7] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod, "CHoG: Compressed histogram of gradients a low bit-rate feature descriptor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Jun. 2009, pp. 2504–2511.

[8] M. Jia, X. Fan, X. Xie, M. Li, and W. Ma, "Photo-to-search: Using camera phones to inquire of the surrounding world," in *Proc. Mobile Data Manag.*, 2006.

[9] S. Tsai, H. Chen, D. Chen, G. Schroth, R. Grzeszczuk, and B. Girod, "Mobile visual search on printed documents using text and low bitrate features," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2601–2604.

[10] X. Liu, Y. Lou, A.W. Yu, and B. Lang, "Search by mobile image based on visual and spatial consistency," in *Proc. IEEE Int. Conf.Multimedia Expo*, 2011, pp. 1–6.

[11] 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 search,” in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 3885–3888.
[12] 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.
[13] S. S. Tsai, D. Chen, H. Chen, C.-H. Hsu, K.-H. Kim, J. P. Singh, and B. Girod, “Combining image and text features: a hybrid approach to mobile book spine recognition,” in *Proc. ACM Multimedia*, 2011, pp. 1029–1032.

