Australian Journal of Basic and Applied Sciences, 8(10) July 2014, Pages: 7-11

AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

An Efficient Interactive Mobile Visual Search Using Multipart Region based Matching (MRM) Algorithm

[1] P. Sivakumar, [2] S. Perumal Sankar, [3]M. Praksh, [4]S. Malathi, [5]A. Duraisamy, [6]S. Muthukumar

[1]*Assistant Professor, K.S.R College of Engineering, Tamilnadu,Computer Science and Engineering,Namakkal,Tamilnadu , India.*
[2] *Professor, CAPE Institute of Technology,Kanyakumari Dt,Tamil Nadu, India.*
[3]*Assistant Professor, K.S.R College of Engineering, Tamilnadu,Computer Science and Engineering,Namakkal,Tamilnadu, India.*
[4]*Final Year,M.Tech Multimedia, K.S.R College of Engineering, Tamilnadu,Computer Science and Engineering,Tamilnadu, India.*
[5]*Assistant Professor, Department of Information Technology, University College of Engineering Tindivanam, Tamil Nadu, India.*
[6] *Professor , Department of CSE, National Institute of Technology, Puducherry, Karaikal, India.*

ARTICLE INFO

Article history:
Received 25 April 2014
Received in revised form
8 May 2014
Accepted 20 May 2014
Available online 17 June 2014

Keywords:
Mobile visual search, Exemplary image, Multimodal input, Image composition, Interactive search, Mobile device.

ABSTRACT

Background: Searching is becoming pervasive and one of the most popular applications on mobile devices. Text, voice, sketches are used as query to search images on mobile devices. A multimodal interactive image search on mobile device is proposed to visual search which focuses on image composition instead of typical image retrieval. Search performance is the main constraint for any kind of search techniques. **Objective:** This system introduces “Joint search with Image, Speech, And World Plus” (JIGSAW+) for multimodal input and user interactions with mobile devices. Users may have pictures or images in their mind but have no idea to describe or names to address them. It can be resolved by speech. **Results:** The speech of the user is recognized as query by interactively composing a visual query using exemplary images. By using this user can easily find the desired images through a few interactions with mobile devices. The JIGSAW+ system is implemented in three modules. First the images are represented based on features of the images known as segment based image representation. Secondly relative position checking is used for finding images. Finally, inverted index are constructed for matching this system shows the user desired image is found easily and the search performance is improved significantly **Conclusion:** We have proposed an interactive mobile visual search system which allows the users to formulate their search through multimodal interactions with mobile devices. Mobile visual search takes the advantages of multimodal and multi touch functionalities on the phone. The proposed system provides a game-like interactive 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 game-like image search system.

To Cite This Article: P. Sivakumar, S. Perumal Sankar**,** M. Prakash, S. Malathi, A. Duraisamy, S. Muthukumar., An Efficient Interactive Mobile Visual Search Using Multi-part Region-based Matching (MRM) Algorithm. ***Aust. J. Basic & Appl. Sci.,*** *8(10): 7-11, 2014*

INTRODUCTION

Image search is a hot topic in both computer vision and information retrieval with many applications. The desktop image search systems with text queries have dominated the user behavior for a long period. More consumers use phones or other mobile devices as their personal concierges surfing on the Internet. Along this trend, searching is becoming pervasive and one of the most popular applications on mobile devices .It is reported that one-third of the Internet search queries will come from phones in future. The bursting of mobile users puts forward the new requests for image retrieval (Lin Chen, A., Jin Liu 1a). The images are searched based on the query given by the user. Text, Voice, Sketches, Photo and Content of the images are used as query to search images on mobile devices. In the text-based search, the user can type an entity name to find the images. The photo-to-search is becoming pervasive as the development of the computer vision. This enables the user to capture photos using the in-built camera on the phone and then initiate search queries about objects in visual proximity to the user. Then voices are used as query to search images. Even though voice queries are available on some devices, there are still many cases that semantic and visual intent can hardly be expressed by these descriptions for search. The content-based image search, can accept single images as search queries, and return to the user similar images. Another kind of image search is sketch-based image search (Vijay

Corresponding Author: Dr.P. Sivakumar, Assistant Professor, K.S.R College of Engineering, Tamilnadu,Computer Science and Engineering, Tamilnadu, Namakkal, India.
Tel: 919790155384; E-mail: drsivakumar.p@gmail.com

Chandrasekhar, Gabriel Takacs, 2012). It uses 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. The formation/ construction model of visual search results based on text, voice, photo, sketch query is represented in the figure 1.

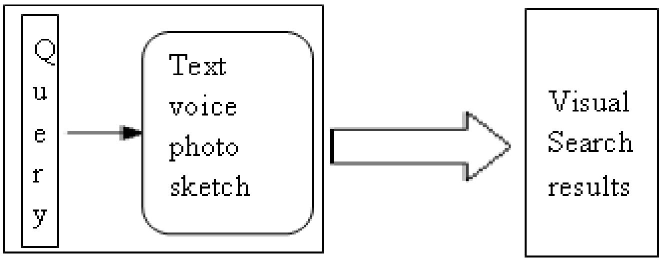

Fig. 1: Mobile visual search.

Image Retrieval:

The image retrieval is based on the input query. Based on the query images are retrieval from database through mobile devices. A small screen limits the presentation of searching results, which requires the top results to be more relevant while on the phone. Using text as query can hardly meet this end. The surrounding texts of the web images are not always correct. The user must know the exact terms the annotator used in order to be able to retrieve the images they want. The image search applications do not perfectly accommodate to the mobile.

Related Work:

The most related works include many multimedia search applications available on mobile devices. Text-based search engines like Google and Bing are still available on mobile devices. But it is neither user-friendly on phone, nor machine-friendly for search engine. Only 2.6 terms of mobile users use for search (Vijay Chandrasekhar, Gabriel Takacs), which can hardly express their search intent. As the pervasive mobile phones support multimodal input such as the built-in high resolution camera, voice, multitouch function, etc for more convenient to user. As the speech recognition became grown-up, phone applications using speech recognition rapidly grows. The most representative application is Apple Siri (MATRIOSKA, 2013), which combines speech recognition, natural language understanding and knowledge-based searching techniques. 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. Google Goggles, Point and Find (Neven, Sr., Hartmut) and SnapTell (Tao Chen, Ming-Ming Cheng) are good examples for this search. Partial duplicate images in their database are used as query for searching images. Traditional features such as MPEG-7 image signature, SIFT, Speeded Up Robust Feature (SURF) (Eitz, M., Fak, IV-Elektrotechnik) and Oriented FAST and Rotated BRIEF (ORB) (http://pointandfind.nokia.com) 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 (Syst, Beijing, China Mingjing Li). Various systems with optimized technique and better indexing are developed to search for landmarks (Lin Chen, A., Jin Liu 1a), books, CD covers etc. Other techniques are also used in visual search such as barcodes and OCR. The proposed JIGSAW+ system represents a new visual search mode by which the mobile users can naturally formulate visual queries to search for images on the go.

The Jigsaw:

The term JIGSAW is a “Joint search with ImaGe, Speech, And Word” (JIGSAW), used for searching images on mobile devices. In the implementation of JIGSAW system, images in the database are divided into 9 x 9 blocks which is the same as Concept Map, but a different matching scheme and different features were used. From each block, color, gradient and local features are extracted into a histogram offline. According to the coverage of an exemplary image, histograms from corresponding blocks are combined into a single feature vector and compared with that of the exemplary image. The position of the exemplar will be slightly shifted to adjacent places to tolerance the position uncertainty. Then the max score is pooled for this single exemplar. Different from common similar image search, a fusion step of several concepts are essential to ensure there exist all those exemplars in the final results. Images related to any keyword are considered in the matching step. Finally, these candidate images are ranked according to their relevance scores and displayed on the phone.

The Jigsaw:

A multimodal interactive image search on mobile device is proposed to visual search (Houqiang, Li, Yang Wang) which focuses on image composition. The term JIGSAW+ is a "Joint search with ImaGe, Speech, and And Word Plus", for multimodal input and user interactions with mobile devices. With rich interactions and visual techniques, the JIGSAW+ enables the user to conduct a image search with visual aids. The objective is to design a efficient visual aided image search application on mobile phone. The search procedure of JIGSAW+ is as follows and the figure 2 shows that:

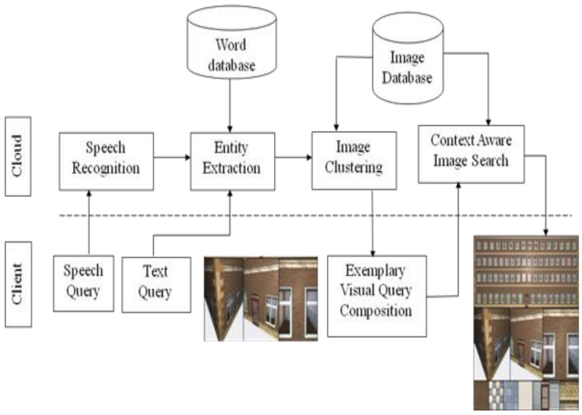

Fig. 2: Visual Search System (JIGSAW+).

1) The user speaks a natural sentence to describe the images, 2) the speech is recognized and then decomposed into keyword(s) which can be represented by exemplary images, 3) the user selects preferred exemplar(s) and then composes image, 4) the composite image is then used as a visual query to search for similar images. Compared with JIGSAW, JIGSAW+ the algorithm has been improved in three aspects: 1) Segmentation-based image representation; 2) relative position checking and 3) inverted index is constructed for matching.

Segment Based Image Representation:

The images are retrieval based on features of images such as color, texture and more. Uniform grid partitions are used to break the original image into smaller pieces of image.

Color Feature Extraction:

Before color feature extraction, the images are over-segmented. The over-segmentation methods use graph-based algorithms that segment an image into many homogeneous regions. Each node in the graph stands for a pixel in the image, with undirected edges connecting its adjacent pixels in the image. The weight of each edge between two pixels reflects their similarity. The similar pixels are merged. Moreover, the similarity of RGB-color space is used instead of gray level, so that inside each piece the color is close.

Texture Feature Extraction:

The most widely used local feature of SIFT (http://pointandfind.nokia.com) to capture local texture patterns in the image. To reduce computing and noise, only prominent regions are used instead of homogeneous regions, prominent regions are detected using Difference of Gaussian (DoG) detector.

Relative Position Checking:

The position of each exemplar in image should be consistent with the composite visual query. The existence scores for all exemplars are obtained for Image I,

$$s = \sqrt[Q]{\prod_{q=1}^{Q} S_q}$$

Eq.(1)

The position of occurrence of each exemplar is used to check the spatial consistency (Liu, X., Y. Lou, 2011). The position of the object can be estimated by

$$\boldsymbol{y} = \sum_{n=1}^{N} \frac{s_n \boldsymbol{y}_n}{S_q}.$$

Eq.(2)

Where y = object position, Sn = components in ingredient vector. The distance between exemplar position x and object position y is used to obtain a measure of spatial similarity

$$p = \exp\{-||x-y||2\} \quad (3)$$

If the object occurs in the identical position as the user demand, this spatial similarity will get the maximum value of 1, which means the position is exact the same. If the object is far from the demanded location, the spatial similarity is small. For a pair of entities i and j, the spatial consistency is obtained by

$$Pi, j = \cos(xi - xj, yi - yj) \quad (4)$$

Finally, the visual similarity between the image and visual query is combined by:

$$Simv(Q, I) = p.s \quad (5)$$

This similarity reflects both the existence and the spatial consistency of all the exemplary images.

Inverted Indexing:

The tags are used for indexing method. The stemmed tags are indexed into an inverted-file index with tf-idf (term frequency-inverse document frequency) weights, so that another similarity between image and query from text information can be obtained. The similarity between Q and I can be calculated as follows:

$$Sim(Q, I) = Sim(Q, I).Sign(Simt(Q, I)) \quad (6)$$

Where sign(x) is a binary function which is 1 if x>0, otherwise 0.
Sim(Q, I) to rank images in our database and show the top-ranked images in the result page.

Algorithm:

Algorithm 1 The multi-part region-based matching algorithm:

1: Retrieve the similarity for each exemplary image in the reverted file;
2: Estimate the position by averaging the positions of matched visual words.
2: For multi-exemplar entities, find the max similarity.
3: Merge the similarities from different entities by (1).
4: Obtain the spacial consistency score by (3) and (4).
5: Combine the merged similarity and position consistency.

Experiment Analysis:

This experiment is built on the mobile phone or we use a system with phone application for testing the multimodal visual search system. It needs server with 32 GB memory and 2.0 GHz 16-core CPU, and a front-end interface application is developed on a Windows Phone 7 device. The topmost text box accepts the text queries input by either typing or speech. The record button is used to record the speech. It is used only when the text box is empty. The speech is then sent to an online service and the recognized text will appear in the text box. Then the keywords are extracted and listed below the text box as tags. When the user selects one of the tags, a list of exemplary images representing this tag will appear below. The user can drag any of the exemplary images down onto the canvas area below. On the canvas, the user can reposition and resize the exemplars. Once the user finishes the visual query, tapping the search button can launch the visual search. The search results will be listed in a new page.

Performance Analysis:

To evaluate the performance of the proposed search algorithm, it is deployed on a back-end server. In the experiments, we compared the performance of text-search, MindFinder, Concept Map, JIGSAW, and JIGSAW+. For all methods, images from a web image search engine are used as exemplary images and Flickr images are used in the database. JIGSAW+ outperforms all the other four methods because of its three advantages. First, it has a much natural representation of features, whereas the block-based representation and indexing are artificial and unnatural. Second, it considers relative layout of the exemplars. The retrieval schemes of all the other methods failed to break the absolute position constrain. Third, all of the other methods have fatal defects. MindFinder has no fusion scheme when there are multiple objects. If only one of the exemplar matches well and offers high relevance, it will be ranked in top even though other objects are missing. Concept Map and JIGSAW cannot index images in the database, only top images of the text results are checked for the visual similarity, otherwise it will be very slow. Moreover, both Concept Map and JIGSAW are not able to index and retrieval large scale database. However, JIGSAW + has no such problems. Even in some cases JIGSAW+ can have a moderate performance without text. JIGSAW+ performs even better because it grasps all the meaningful information transformed from the user's visual intent. The JIGSAW keeps only global information, whereas in JIGSAW+, the region-based method considers both global and local (position) characteristics.

Conclusion:

Thus an interactive mobile visual search system which allows the users to formulate their search through multimodal interactions with mobile devices. Mobile visual search takes the advantages of multimodal and multitouch functionalities on the phone. The proposed system provides a game-like interactive 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 game-like image search system.

REFERENCES

Dima Kassab and Xiaojun Yuan, 2013. College of Computing and Information, University at Albany, State University of New York and Albany NY, USA. Understanding the information needs and search behaviour of mobile users, 17-4.

Eitz, M., Fak, IV-Elektrotechnik & Inf., Tech. Univ. Berlin, Berlin, Germany K. Hildebrand, T. Boubekeur, M. Alexa, 2011. Sketch-based image retrieval: Benchmark and bag-of-features descriptors.

Feng Jing State Key Lab. of Intelligent Technol. & Syst, Beijing, China Mingjing Li; Hong-Jiang Zhang;BoZhang, 2004. An efficient and effective region-based image retrieval framework.

Felzenszwalb and D. Huttenlocher, 2004. Efficient graph-based image segmentation, Int. J. Comput. Vis., 59(2): 167-181.

Houqiang, Li, Yang Wang, Tao Mei, Jingdong Wang and Shipeng Li, 2013. Interactive Multimodal Visual Search on Mobile Device.

Lin Chen, A., Jin Liu 1a, Liang Cao, a School of Geodesy and Geomatics, Wuhan University, Wuhan, 430079 China."Image Matching by Affine Speed-Up RobustFeatures".

Liu, X., Y. Lou, A.W. Yu and B. Lang, 2011. Search by mobile image based on visual and spatial consistency, in Proc. IEEE Int. Conf.Multimedia Expo.

Neven, Sr., Hartmut (Aachen, DE) Neven, Hartmut (Malibu, CA, US) United States Patent 7565139, Image-based search engine for mobile phones with camera, www.freepatents online .com.

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

MATRIOSKA: A Multi-level Approach to Fast Tracking by Learning, 2013. 17th International Conference, Naples, Italy, 9-13.

Tao Chen, Ming-Ming Cheng, Shi-Min Hu-Tsinghua University, Ping Tan-National University of Singapore, Ariel Shamir,2009,"Sketch2photo: Internet image montage", The InterdisciplinaryCenter.

Vijay Chandrasekhar, Gabriel Takacs, David M. Chen, Sam S. Tsai, Yuriy Reznik, Radek Grzeszczuk, Bernd Girod, 2012. Compressed Histogram of Gradients: A Low-Bitrate Descriptor.

Ystad, Sweden, 2011. "Mobile Visual Search from Dynamic Image Databases", 17th Scandinavian Conference, SCIA2011.