M CBIR Interactive Multimodal Visual Search on Mobile Device with Image Composition + LAB-ANN

Yojana S. Pawale, Research Scholar, Deptt. of Information Technology, Bharati Vidyapeeth Deemed University, College of Engineering, Pune City, India
Prof. Prakash Devale, Professor, Deptt. of Information Technology, Bharati Vidyapeeth Deemed University, College of Engineering, Pune City, India

ABSTRACT

This paper proposes a technique - M CBIR Interactive Multimodal Visual Search On Mobile Device With Image Composition + LAB-ANN. In this system, images can be searched on mobile devices by taking multimodal input i.e. a voice, image and text to the system. The system performs a joint search with image, speech and text to retrieve desired images from it. The best thing of it is that the system also makes a use of ANN technique to retrieve more relevant images from it. This technique is useful in a case where users do not know exact name of an image but have a picture in their mind then by describing it using speech or text user can easily find the desired images on their mobile phones. This technique is also useful for users who want more relevant images of a particular image by providing that image as an input query. This system is more powerful than existing system and capable of retrieving more relevant images of a given input by applying text composition+ ANN technique. This system first make a composed image of a given input and then again use that image as an input, it tries to search for more relevant images in their database by first applying an algorithm and then ANN technique. Hence the proposed system is able to achieve 5% gain in terms of search performance and is ten times faster.

Keywords
Image Search, ANN, LAB image, Image Processing, Multimodal Input

1. INTRODUCTION

Image search is a hot topic in both computer vision and information retrieval with many applications. It is more difficult to search and find desired images on mobile phones rather than desktop computers because of following reasons:

First, is user interface where on desktop computers, they accept an inputs through keyboard and mouse but in case of mobile phones, they always provides input through cameras, GPS, microphones and multi touch screens. Hence it is difficult to give an input on mobile devices and also not user-friendly.

Second reason is user’s search interest. Most of the time, mobile users search an image to find local information such as local spots and local map. Hence they have to either type an entity name or lookup on online local map to find desired images. The third reason is small screens. The small screen of mobile devices affects the presentation of searching results. Lastly, Map search, photo to search is not much popular on mobile devices compared with text search because the existing image search applications are not perfectly accommodate to the mobile and local oriented user intent. Hence it is rarely used on mobile devices.

By considering above issues of image searching on mobile devices, we have introduced a new approach- M CBIR Interactive Multimodal Visual Search On Mobile Device With Image Composition + LAB-ANN. This technique accepts a multimodal input as voice, text, image to find desired images. The system performs image searching and retrieving by using image composition with LAB-ANN (Artificial Neural network) technique. This system is useful in a case where users does not know exact name of an image to be search. Then by describing it using either text or speech or any other relevant image, they can easily find more relevant images. The best part of this system is that it uses ANN technique to find exact targeted images by considering user’s psychological factors behind image searching. This system provides faster recognition mechanism through the use of ANN technique. Hence it helps to improve the performance of the system. System also make a use of LAB technique that exactly capture brightness, color factors of an input image which will be helpful to perform further image processing. Hence it improves the performance of image search and is ten time faster than other ones.

2. ANALYSIS OF EXISTING SYSTEMS

The basic technique for image searching is text-based search, where input is given as a text to system and finds desired images from it. Traditional search engines like
Google and Bing are still use this technique. But the fact is lengthy text queries are neither user-friendly on phone, nor machine friendly for search engine because the mobile users use only 2.6 terms on average or Search[2], which is hardly express their search intent. As a technology improved, Speech recognition technique is popularized in phone applications. The most popularized application is Apple Siri[3], that provides knowledge-based searching techniques with the help of speech recognition, natural language understanding. In this technique, user can able to give an input as a speech and get information and knowledge from it on mobile devices.

Another mechanism is popularized for image search is photo-to-search application in which users are able to search for what they see by taking a photo of it and gives an input to the system. For this technique, Google Googles[4], Point and Find[5], Snaptell[6] are the good examples in this field. These applications are not only search for the exact partial duplicate images in their database but also provide related information of the query images. But these applications provides the searching mechanism only for some vertical domains like-products, landmarks, CD covers etc... where the partial duplicate images of the query image have been indexed in their database.

Now, researchers for mobile search also focus on mainly to photo-to-search technique. Their research topic describes visual feature design, database indexing, and transmission. Traditional features such as MPEG-7 image signature, SIFT, Speeded Up Robust Feature(SURF)[7] and Oriented FAST and Rotated BRIEF(ORF)[8] 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) [9]. It can quantize and encode gradient histogram by making a use of Huffman and Gagie trees and produce very low bit-rate descriptors. Various systems are also developed to search for landmarks [10], books [11], CD covers [12] etc. with the help of optimization and better indexing technique. In [13], indexing with bundle of features is introduced in a poster and CD cover retrieval system for mobile visual search. In [12], CD cover search system compares different local descriptors where SIFT is considered as the best performed feature and CHoG provides so many advantages in low-bit transmission. In [14], Bernd Girod et al. gave a comprehensive overview of photo-to-search from the architecture of an efficient mobile system to the framework of an image recognition algorithm. Some other techniques are also developed for visual search such as barcodes and OCR [15], [16].

The recent system for image search is JIGSAW+ [1] is popularized. In this technique, system accepts an input as a text, image and speech. Then converts it into keywords and find relevant images for it. After that collect these images and represents exemplary images of these by applying mining and clustering technique. Many approaches are also available in this step like visual query suggestions [17]. And lastly, user select relevant exemplary image and gives as a input to the system where on the other hand, system accept that image and through resizing, positioning input image, provides relevant output images to the user. Recently, this technique is a powerful technique for image search on mobile but it missed the issue of artificial intelligence. This system does not consider an issue of user’s intention to be search. For example, If user gives a query as a ‘head; then system is not focused on the point as whether user wants images related to ‘human head’ or ‘Head-of-department’.

Hence by considering all above issues of image searching on mobile device, We have proposed a new technique- Mobile CBIR Interactive Multimodal Visual Search On Mobile Device With Image Composition + LAB-ANN. The functionality of this system is mostly same as JISAW+ technique but additionally applies LAB and ANN approach to find more relevant images and make a faster recognition mechanism.

3. PROPOSED SYSTEM

The main objective of proposed system is to design an efficient image search application on mobile phone through the use of LAB and ANN technique. The fig.1 (a) and fig.1(b) shows the comparison between two systems and fig.1(c) shows the basic structure of our system.

As shown in fig.1 (c), the system combines functionality of systems shown in fig. 1 (a) and fig. 1 (b). The system accepts an input as either voice or text or image. If it accepts an input as voice then converts into text and search for relevant images. After that one composed image is provided to the user where user selects one relevant composed image from it and gives it to the system as a query. Then, system again searches for more similar images of given query image.
Fig. 1. Three modes of mobile visual search: (a) voice/text-to-search, (b) photo-to-search, (c) our proposed visual search system.

Fig. 2. Architecture of our proposed system.
4. ARCHITECTURE OF PROPOSED SYSTEM

The overview of proposed system is shown in fig.2. The system consists of following phases:
1) The user gives an input query as a voice or image or text to the system. 2) If an input is a speech or voice then it converts it into text using Android’s TTS API. 3) Text is further decomposed into keywords. 4) According to keywords, composed image is provided to the user. 5) User selects one relevant composed image from it and gives it to system as a query. 6) Then, system applies some image processing steps on query image such as converts RGB color image into LAB form then converts into graphical form as a histogram. After that image is quantized and then normalized to apply ANN technique on it. 7) ANN technique is applied on normalized image. For that BPNN mechanism is applied on image. 8) In the last step, ANN classification is performed, where system searches an image in its own database by taking ANN features from database and result of BPNN process. Then by comparing these parameters, system finds more relevant images from database and provides to the user as a final result.

5. CONCLUSION

The system proposed a method as a multimodal approach for image searching on mobile devices. This technique is useful in cases where user does not know the exact name of an image but having some pictures in their mind to describe it. Then, by giving an input as a speech or text, user can find desired image from the system. This is also useful in cases where user can provide input as an image and get more number of relevant images of the same one. This technique converts input into tokens and finds relevant images of each token and make composite image. Then use that image as an input to the system, it applies some image processing technique and lastly, converts it into normalized form. Then in the last step, before search into database for finding relevant images, system applies ANN technique and divides images into two parts—positive and negative. Lastly, positive set is provided to the user as a final result.

REFERENCES


