ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May2015

I2MMS: An Interactive Multi Modal Visual Search Technique

Yojana S. Pawale[1], P. R. Devale[2]

Research Scholar, Dept. of Information Technology, B. V. D. U, College of Engineering, Pune City, India [1]

Professor, Dept. of Information Technology, B. V. D. U., College of Engineering, Pune City, India [2]

Abstract: Now days, image searching technique is popularized on mobile phones where mobile users search and find desired images through their mobiles. But it is not easy to find desired image through mobile, because mobile phone' screen is too small to display full image. It is also difficult for users to provide input on mobile phones to find targeted image. Hence, by considering all these issues, we have developed one image searching technique called – I2MMS: An Interactive Multi Modal Visual Search Technique. In text based technique, if user provides lengthy queries to find desired image then it will be difficult for system to process that whole query and provides desired images to users and also not user friendly. Hence, we have proposed a new technique for image searching where system accepts multimodal input that can be voice, text and image. This system is useful in a case where users do not know exact name of an image but by describing it using either text or speech or by providing any other relevant image, users can easily find targeted image. The ANN technique is also added into it to increase the performance of the system. The system works in three phases- 1) Image Composition, 2) Image Processing, 3) ANN. Lastly, before providing images to users as an output, images are divided into two sets- positive and negative. All relevant images are present in positive set and non relevant images are present in negative set. Lastly images present in positive set are provided to the user as a final output. Hence our approach improves the quality of image searching technique and provides easy way to finds targeted image.

Keywords: Image Processing, Artificial Intelligence, Content Based Image Retrieval, Multimodal Input

I. INTRODUCTION

Image searching is a technique of searching, finding and retrieving desired images from database. A large database is present at server site that contains no of images. The system accepts input query from user and provides a set of number of relevant images to user according to user's query. Currently, image searching technique is populated on mobile phones. Most of the time, image search is performed by mobile users to find local information like local maps. But it is not easy to search and finds targeted images through mobile phones. It may face many problems. For example, Mobile phone's screen is too small to display full length image. Hence, it may affects on presentation part of the system. Another drawback is - a way of providing input to the system. On computer system, keyboards and mice are used to provide input mice but in case of mobile phones, input is provided by cameras, GPS, microphones and multi touch screens. Hence, it is very difficult to provide inputs to system and not user friendly as well as machine friendly. Many techniques are developed for image searching. The basic technique is developed for image searching is text-based technique in which user gives text query to system and retrieves desired images. But the drawback of tit is that if user gives lengthy query to find an image then that lengthy query is neither user friendly nor machine friendly because the fact is that mobile users use only 2.6 terms on average for search [2], which can be hardly express their search intent. One more technique is developed for image searching is photo – to – search technique. In this technique, user provides a query to system as an image to find some more relevant images. Google Googgles[4], Point and Find[5], Snaptell[6] uses this technique. Some extra information of an output images is also provided by these applications. The limitation of these applications is that it provides searching mechanism only for landmarks, CD covers, Products etc. One technique is popularized on mobile phone is speech recognition technique. Apple Siri[3] is an application that uses this technique in which speech recognition and language understanding approaches are used to provides knowledge based searching mechanism. Recently one more technique is developed for image searching is JIGSAW+ [1]. This system performs join image search by accepting input as text, image or speech. This technique is also popularized on mobile phones but it does not consider user's intention to be search. It missed an issue of artificial intelligence. For example, if user gives text query as an 'Apple'; then system does not focus on the point that whether user wants images on 'apple fruit' or 'apple company's products'.

Hence by referring all these techniques, we have developed one mechanism called I2MMS: An Interactive Multi Modal Visual Search Technique. The system is multimodal where text based and image based techniques are combined with speech based technique to retrieve quality images by adding ANN mechanism in it. This mechanism provides multimodal approach where system accepts input as an image, text or speech. This system is mostly same as JIGSAW+ but also applies LAB and ANN technique to improve the performance of the system and provides faster searching mechanism to users. The system

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May2015

works in three phases as – 1) Image composition 2) Image processing 3) ANN. The plan for remaining sections is as follow- section 2 describes concept of proposed system. Section 3 represents system's architecture. Section 4 discussed detailed description of each module of proposed system. Section 5 shows experimental work of proposed system and lastly section 5 concluded this concept.

II. PROPOSED SYSTEM

The proposed system provides joint mechanism for image searching on mobile phones by combining text based and image based searching mechanisms with speech recognition techniques. The system is multimodal which accepts input as an either image or text or speech.

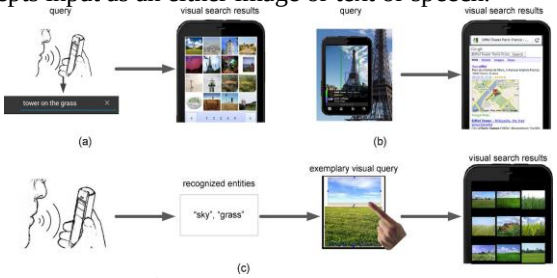

Fig.1. Three modes of mobile visual search: (a) voice/text-to-search, (b) photo-to-search, (c) our proposed visual search system.

Above figure shows the basic idea of proposed system. Fig. 1(a) represents voice/text-to-search technique in which user provides input as either voice or text and retrieves relevant images. Fig. 1(b) represents photo-to-search technique where user gives image query and retrieves more relevant images. Fig. 1(c) represents basic idea of proposed system. In proposed system, user provides input query to the system that can be either text or voice or image and then according to that query, composite images are provided to the users. After that User selects one image among them and retrieves more relevant images. For example, if user wants to search an images for Punjabi restaurants in pune having red door. Then, system searches for composite images in their database having all these entities as - Punjabi restaurants, pune and red door. Then user selects one composite image among them and retrieves more relevant images.

III. ARCHITECTURE OF PROPOSED SYSTEM

The system works in three main phases – 1) Image Composition, 2) Image Processing, 2) ANN.

As shown in fig. 2, the system accepts input as either voice or text or image. If system accepts input as a voice then converts it into text and searches for no of relevant composed images within its own database and shown back to the user. Then user selects one composed image among them. After that system again accepts this image query as an input for further processing. User can directly provide image query at the first time for retrieving more relevant images. In the next phase image processing is performed on query image to extract features of it and provides to the next phase which requires for applying ANN technique. Before processing the query image, it converts into LAB form to improve the quality of an input image.

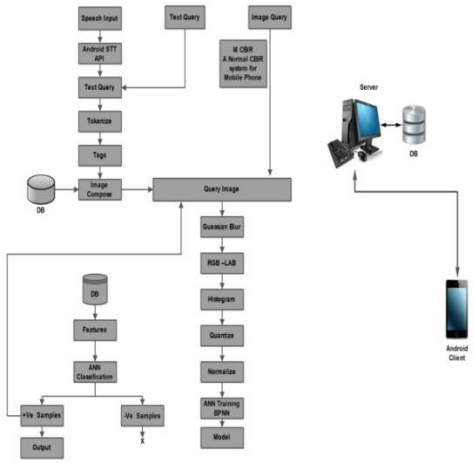

Fig.2 Architecture of our proposed system

In the last phase, the result of image processing is used to apply ANN mechanism on it. Here, BPNN mechanism is applied on received image for further processing. Then, system searches an image in their own database by taking ANN features of it and result of BPNN process of an input image and compares parameters of them and provides most relevant images to users. Before providing relevant images to user, two sets are created by the system-positive and negative sets. Positive set contains number of relevant images and negative set contains number of non relevant images. Lastly, images present in the positive set are provided to the user as a final output. Here, users can again select one output image and provides it as an input query to the system to find more relevant images.

A. Image Composition

In the first module, user provides input query as an either speech or text or image. If user provides speech query then system converts it into text. Android's STT API is used to convert speech into text. Then, tokenization process is applied on text query where text is decomposed into different tokens. Tokens are converted into tags and passes to the next phase. This step is performed because system searches images within database according to the tokens. Hence, these tokens are separated by tags. After that, System searches for composed image in its own database according to the keywords and provides to the users. For example, if the text query is – "Restaurants in Pune". Then system accepts that text and converts it into keywords as- 'Restaurant' and 'Pune'. These tokens are separated by tags and related composed images are searched in system's database and provides to users. Here, System searches for composed images that having both the keywords i.e. 'Restaurant' and 'Pune'. User selects one composed image among them. Then, system accepts that composed image as an input query and passes it to next phase. It is also possible to provide input as an image at the first time to the system and finds more relevant images from system by performing next process.

B. Image Processing

This phase accepts input image and blurring it by applying Gaussian function. The image is converted into Gaussian form to reduce image noise. Then image is converted into

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May2015

LAB form to improve the quality of an image. Here, L is stands for lightness and a, b are called color-opponent dimensions, based on nonlinearly compressed coordinates.

Fig. 3(a) Original image

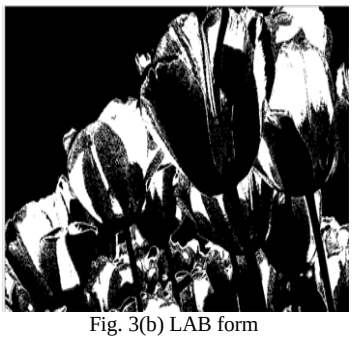

Fig. 3(b) LAB form

Fig. 3 shows LAB form of original image. LAB image is simply an image present in black and white form but in qualitative form. It helps to extract correct features of input image.

Then, histogram is created of LAB image. Histogram is a graphical representation of an image where horizontal axes represent tonal variation and vertical axes represent number of pixels in that particular tone. Histogram helps to judge the entire tonal distribution at a glance. Then, image is quantized after histogram is created. In the last step of image processing, image is normalized to apply ANN technique on it to provide more relevant images to the user.

C. ANN

This is the last phase of the system in which system accepts the result of image processing to extract required features of an image to apply ANN mechanism on it. This phase is called as a key part of the system because it focuses on the user's psychological factors behind searching. It concentrates on user's intention to be search that is what exactly user wants.

For example, if user gives a text query as a 'Head', then the point - whether user wants images related to 'Human head' or 'Head – of – Department' is considered here. For that BPNN mechanism is applied on image and then compares ANN parameters of an images present in their database with the parameters of query image taken form the result of BPNN process and finds most relevant images and provides to the user.

IV. EXPERIMENTAL WORK

Initially, we have provided one form to user in which user enters IP address of server to start an application. Then, on the next form, three options are shown to user to find relevant images as shown in fig. 5.

Fig. 5 Multimodal input

Depending on the user's choice, image is searched and retrieved from database. Fig. 6 is showing some result of our system. Here, user had provided text query to the system. Depending on that query some composed images are retrieved and shown to user as an output.

Fig. 6 Result of text query

If user selects one composed image among these two output images, For example, 'Apple' image. Then system accepts this apple image as an query and provides some more images related to an 'apple' image as shown in fig. 7

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May2015

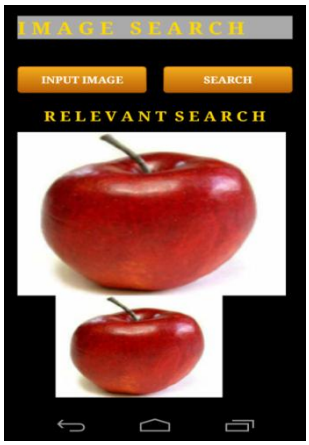

Fig. 7 Relevant images

V. CONCLUSION

The system provides powerful mechanism for image searching on mobile by providing multimodal approach. User can be able to provide any kind of input that is text, image and speech. The system concentrates on user's psychological factor behind image searching. This system works well in a case where user does not know exact name of an image then by describing it using text or speech, user can be able to get desired images. Lastly system separates number of images into two sets as positive and negative. Positive set represents relevant images and negative set represents non relevant images. Positive set is provided to user as a final output.

REFERENCES

[1] Houqiang Li, *Senior Member, IEEE*, Yang Wang, Tao Mei, *Senior Member, IEEE*, Jingdong Wang, *Senior Member, IEEE*, and Shipeng Li, *Fellow, IEEE* "Interactive Multimodal Visual Search on Mobile Device," IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 3, APRIL 2013.

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

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

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

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

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

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] Z. Zha, L. Yang, T. Mei, M. Wang, and Z. Wang, "Visual query suggestion," in *Proc. ACM Multimedia*, 2009, pp. 15–24.