International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

A Survey on Multimodal Visual Search Methods

Seema Ranuji Ghule[1]

Pune Institute of Computer Technology, Savitribai Phule Pune University, India

Abstract: ***Multimodal visual search play vital role in image processing and computer vision. The Visual Search system has input in the form of keywords or visual images and provides information regarding destinations, artworks, books, wine, branded things, product's catalogs, etc. Visual search support features like color, texture and shape. Feature extraction is an important step in multimodal visual search. In this paper different multimodal visual search methods are studied. This paper attempts to provide a comprehensive review of multimodal visual search methods like text based search, image based search and speech based search.***

Keywords: Multimodal visual search, Color histogram

1. Introduction

The mobile industry is going through remarkable change over the past few years with significant advances in areas of communications and multimedia system. 3G services are already within the market and provide a good bandwidth to fulfill the rising demands of the users with delivery of high quality multimedia system services. Currently state-of-the-art multimedia system compliant mobile phones equipped with digital cameras and camcorders have inherent support for network connection and so, change access to large amount of digital media. The processing power and memory capability of mobile phones are increasing all the time. Now a days, mobile platforms provide rich programming API (Application Programming Interface) to developers. With the generation of digital media by capturing and storing facility in smart phones, there's a requirement for content management and system to produce fast retrieval of digital media items from large archives. Therefore, it's become vital to retrieve desired data with efficiency and efficiently using these devices.

Because of the advances in digital photography, storage capability and networks speed, storing large amounts of high quality pictures has been created possible. Digital images are utilized in a wide range of applications like medical, virtual museums, military and security purposes, and personal photo albums. However, users have difficulties in organizing and searching large numbers of images in databases, as the current commercial database systems are designed for text data and not well compatible for digital images. Therefore, an efficient way for image retrieval is desired. A multimodal image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. The Visual search engine aims at serving to users find and quickly get information regarding their things of interest. Information in the form of text or a visual image is effectively accessible over the web .The Visual Search engine takes input in the form of keywords or visual images and provides information about destinations, artworks, books, wine, branded items, product's catalogs, etc

The rest of the paper is organized as following. In section 2 describes text based image search. Section 3 briefly describes the image based retrieval Section 4 describes image searching based on voice or speech will be reported in Section 5, the paper is concluded in section 6.

2. Multimodal Visual Search Architecture

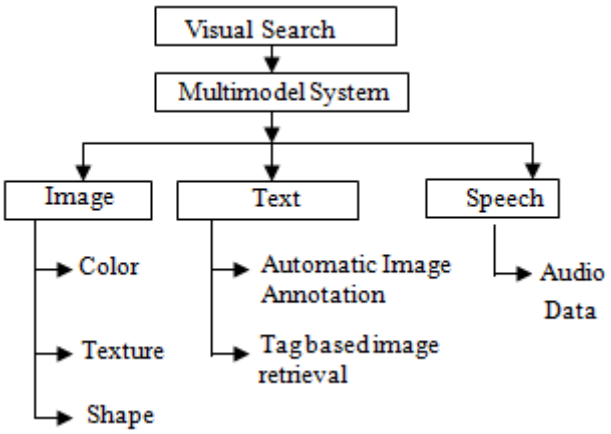

Figure 1: General visual search block diagram

A. Text Based Search

Text based image retrieval is traditional methods. In this method text or keywords are used for describing content of image. By exploiting such rich semantic textual descriptions of Web images, the TBIR has been widely used in popular image search engines like, Google, Bing, Flickr and Yahoo. Specifically, a user is required to input a keyword as a textual query to the retrieval system. Then the system returns the ranked relevant images whose surrounding texts contain the query keyword, and the ranking score is obtained according to some similarity measurements between the query keyword and the textual features of relevant images. However, the retrieval performance can be very poor, particularly when the textual features of the Web images are sparse and noisy in a high-dimensional space. To overcome the defects in the former system, we propose a new framework for the image retrieval system. Following are many techniques for TBIR

1. Automatic Image Annotation.

Automatic image annotation (AIA) is the process by which a computer system automatically assigns metadata in the form of adding captions or keywords or tags to a digital image. Content based ranking of images is harder than for textual documents because they do not have words in that part. Image search techniques are working mainly based on using annotations and semantic tags that are present in the images.

However, tags are entered by the users manually which consumes a large amount of time for the tons and tons of images present in the database. Thus, AIA has been a most challenging task in the past decades.

2. Tag based image retrieval

TBIR is the most efficient technique in image retrieval but it is dependent the tags. The tags are added manually by the users during the time of uploading. To overcome that, automatic annotation technique is applied to add tags automatically to the images during the upload of an image in any database. The tags are the text or annotation present in the image.

Survey on text based image search

Lin Chen [1] proposed tag based image retrieval system. Support vector machine is used for classification. Laplacian regularized least squares method used for refine the relevance scores of test photos by utilizing the visual similarity of the images within the group. Bo Luo[2]they describe hybrid method. System search text only, image only or combination of text or image. There are many images are available on WWW, This system filter that images. So retrieval performance is increased. Lei Wu[3] present algorithm for tag completion. to automatically fill in the missing tags as well as correct noisy tags for given images. Results are verified using two set like automatic image annotation and tag based image retrieval.

B. Image Based Search-

1) Color Feature Extraction Method:

In content based image retrieval system color is important feature. However the robustness, effectiveness, and efficiency of its use in image indexing are still open issues. In image pre-processing, the features used to represent color information and the measures adopted to compute similarity between the features of two images are critically analyzed [4]. Color is invariant to scaling, translation and rotation of an image.

Color Space

Color space consists of three dimensional spaces and color is used as a vector in it. Color Spaces are required for description of color based retrieval of image [5]. Mostly RGB, LAB, LUV, HSV, YCrCb and opponent color space are used. The selection of color space is done from uniformity Characteristics [6] and uniformity means to have colors points having similar distance in color space as perceived by human eye.

Following are color feature extraction methods:

1) Color moments(CM),
2) Color correlogram
3) Color histogram,

Color Moment

Quantization effect in color histogram is reduced in color moment. Stricker and Orengo[7] describe color moment as feature vector for image retrieval. Only consider first moment (mean), the second moment (variance)and the third moment (skewness) are consider for the feature vectors. Euclidean distance is used for similarity between two color moments. The common moments are mean, standard deviation and skewness, these are describe using following calculations

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij} \quad (1)$$

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}(fij - \mu_i)^2\right)^{\frac{1}{2}} \quad (2)$$

$$\gamma_i = \left(\frac{1}{N}\sum_{j=1}^{N}(fij - \mu_i)^3\right)^{1/3} \quad (3)$$

Where fij is the color value of the i-th color component of the j-th image pixel and N is the total number of pixels in the image. μi, σi, γi(i=1, 2, 3) used for calculating the mean, standard deviation and skewness of each channel of an image respectively.

Color Correlogram:

Color correlogram [8] was proposed to represent color distributions of pixels and spatial correlation of pairs of colors. The color of any pixel pair is represented by first and the second dimension of the three-dimensional histogram and their spatial distance described by third dimension.

Color Histogram:

Color histogram is the distribution of color in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges that span the image's color space.

$$h_{A,B,C}(a, b, c) = N.Prob(A=a, B=b, C=c)$$

where A, B and C represent the three color channels (R, G, B or H, S, V) and N is the number of pixels in the image. Color histogram is formed by discretizing the colors within an image and counting the number of pixels of each color.

2) Survey on color based feature extraction methods

Wasim Khan et al. [9] proposed a method for image retrieval using histogram values and texture descriptor analysis of image. Color histogram is used for extract color feature and texture features are extracted using entropy, local range and standard deviation of image.Rishav Chakravarti R et al.[10] proposed algorithm which tests a simple color histogram based search and retrieve algorithm for images. Color histograms does not necessarily allow the relevant images as seen by the algorithm to be the same as the relevant images as seen by a human.Pawandeep Kaur et al.[11] proposed image retrieval system based on color. The goal of this system is find the images most resembling the query. Similarities differences are calculated in many ways.S.R. Kodituwakku et al.[12] proposed many comparison between different color features for image retrieval. Color histograms, color moments and color coherent vector (CCV) are considered for retrieval.Chesti Altaff Hussain et al.[13] represent color histogram based image retrieval. In this paper query image is segmented into two blocks, and each database image is segmented into two blocks and histograms are also generated separately, after that comparison done separately, and considering all local histograms a sorted order of best

suitable images were generated, finally retrieve images with sorted order.Srikanth Redrouthu et al.[14] describe time comparison of various feature extraction of content based image retrieval. Color features are extracted using color histogram and edge detection is used for shape extraction. Performance evaluation measures are precision and recall used. Malini, R. et al[15] provide color perception histogram for image retrieval using similarity measures. HSV color space model is used. Color pixels are extracted and NBS distance is calculated.

3) Texture Feature Extraction Methods

Texture refers to the visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity[16].Structural methods, including morphological operator and adjacency graph, describe texture by identifying structural primitives and their placement rules. They tend to be most effective when applied to textures that are very regular. Statistical methods, including Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura feature, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and wavelet transform, characterize texture by the statistical distribution of the image intensity. Texture features can be extracted in several methods, using statistical, structural, model-based and transform information.

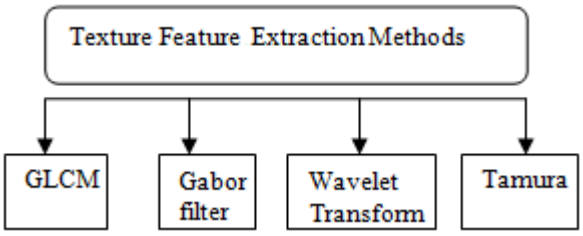

Figure 2: Texture Features Extraction methods

a) *Gray Level Cooccurence Matrix-*

Haralick et al. proposed the co-occurrence matrix representation of texture features [17]. This approach explored the gray level spatial dependence of texture. It first constructed a co-occurrence matrix based on the orientation and distance between image pixels and then extracted meaningful statistics from the matrix as the texture representation.

b) *Tamura Features-*

Tamura et al. explored the texture representation from a different angle [18]. They describe six visual texture properties were coarseness, contrast, directionality, linelikeness, regularity, and roughness.

c) *Gabor Filter Features*

A. K. Jain et al[19] describe unsupervised texture segmentation using Gabor filters. In these filters have been commonly used to retrieve images in the shape of texture.

P. Aigrain et al[20]provide survey of content-based representation and retrieval of visual media. Gabor filters are appropriate to reduce the joint ambiguity issue in space and frequency domains and used as tuned and scaled edge detector .

d) Wavelet Transform Features

Ma and Manjunath [21] evaluated the texture image annotation by various wavelet transform representations, including orthogonal and bi-orthogonal wavelet transforms, the tree-structured wavelet transform, and the Gabor wavelet transform.

4) Shape Feature Extraction Methods

Shape may be defined as the characteristic surface configuration of an object; an outline or contour. It permits an object to be distinguished from its surroundings by its outline. A good shape representation feature for an object should be invariant to translation, rotation and scaling. Shape representations can be generally divided into two categories: boundary based and region based. The most successful representatives for these two categories are Fourier descriptor and moment invariants.

Table 1: Margin specifications

L. Yang et al[22]	fast method of computing moments in binary images
Rui et al.[23]	modified Fourier descriptor which is both robust to noise and invariant to geometric transformations
W. Niblack et al.[24]	chain codes, characteristic, circumference, area and circular degree.

a) Image searching based on Voice or Speech-

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. For example, in a common image search task, the user might have already conceived of the general idea of expected pictures such as color configurations and compositions. However, the users usually have to pick up ideal images amidst much more irrelevant results. In such cases where irrelevant images spoil the results and ruin the user experience, visual-aided tools can largely boost the relevance of search results and the user experience.

3. Conclusion

A wide variety of researches have been made on image search. Every work has its own technique. As a review paper, it might not include each and every aspect of individual works; however this paper attempts to deal with a detailed review of multimodal visual search methods. And we tried to give a constructive idea for future work in this field.

References

[1] Lin Chen, Dong Xu, Tsang, I.W., Jiebo Luo, "TagBased Image Retrieval Improved by Augmented Features and Group-Based Refinement", *IEEE Transactions on Multimedia*, Volume:14, Issue: 4,

[2] Bo Luo, Xiaogang Wang, and Xiaoou Tang, "A World Wide Web Based Image Search Engine Using Text and Image Content Features", *SPIE*, Vol. 5018, 2003.

[3] Lei Wu, Rong Jin, Anil K. Jain, "Tag Completion for Image Retrieval," *IEEE Transaction On Pattern Analysis and Machine Intelligent*, January, 2011.
[4] Raimondo Schettini, Gianluigiciocca, Silvia Zuffi, "A survey of methods for colour image indexing and retrieval in Image database".
[5] W Hsu, T.S. Chua and H. K. Pung, "An integrated color spatial approach to content-based image retrieval, "*ACM Multimedia Conference*, 305-313, 1995.
[6] E. Mathias, "Comparing the influence of color spaces and metrics in content-based image retrieval, " Proceedings of *International Symposium on Computer Graphics, Image Processing and Vision,* 371-378, 1998
[7] M.Stricker and M .Orengo. "Similarity of Color Images, " in Proc. *SPIE Storage and Retrieval for Still Image and Video Databases III*, , San Jose, CA, USA, pp. 381-392, 1995.
[8] J. Huang, "Image Indexing using Color Correlogram, " *IEEE Int. Conf. on Computer Visionand Pattern Recognition*, pp. 762-768, Puerto Rico, June 1997.
[9] Wasim Khan, Shiv Kumar, Neetesh Gupta, Nilofar Khan, "A Proposed Method for Image Retrieval using Histogram values and Texture Descriptor Analysis, " *International Journal of Soft Computing and Engineering (IJSCE),* ISSN: 2231-2307, Volume-I Issue-II, May 2011.
[10] Rishav Chakravarti, Xiannong Meng, "A Study of Color Histogram Based Image Retrieval," *Sixth International Conference on Information Technology: New Generations*, 2009.
[11] Pawandeep Kaur, Sakshi Thakral, Mandeep Singh, "Color Based Image Retrieval System," *IOSR Journal of Computer Engineering (IOSRJCE),* 2012.
[12] S.R. Kodituwakku, S.Selvarajah, "Comparison of Color Features for Image Retrieval," *Indian Journal of Computer Science and Engineering, Vol. 1 No. 3 207-211.*
[13] Chesti Altaff Hussain, Dr. D. Venkata Rao, T. Praveen, "Color Histogram based image retrieval," *International Journal of Advanced Engineering Technology,* E-ISSN 0976-3945, 2013.
[14] Srikanth Redrouthu1, Annapurani.K, "Time Comparison of Various Feature Extraction of Content Based Image Retrieval", *International Journal of Computer Science and Information Technologies*, Vol. 5 (2), 2518-2523, 2014.
[15] R.Malini, C.Vasanthanayaki, "color perception histogram for image retrieval using multiple similarity measures,"*Journal of Computer Science*, 10 (6): 985-994, 2014.
[16] J. R. Smith and S.-F. Chang, "Automated binary texture feature sets for image retrieval, " in Proc. *ICASSP*-96, Atlanta, GA, 1996.
[17] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. On Sys. Man. and Cyb,* SMC-3(6), 1973.
[18] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. On Sys., Man. and Cyb.* SMC-8(6), 1978.
[19] A. K. Jain, and F. Farroknia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, 24(12), 14-19, 1990.
[20] P. Aigrain, H. Zhang, and D. Petkovic, "Content-Based Representation and Retrieval of Visual Media: A State of the Art Review,"*Multimedia Tools and Applications*, 3, 179-202, 1996.
[21] W. Y. Ma and B. S. Manjunath, "A comparison of wavelet transform features for texture image annotation, " *in Proc. IEEE Int. Conf. on Image Proc*., 1995.
[22] L. Yang and F. Algregtsen, "Fast computation of invariant geometric moments: A new method giving correct results," in Proc. *IEEE Int. Conf. on Image Proc.*, 1994.
[23] Y. Rui, A. C. She, and T. S. Huang, "Modified fourier descriptors for shape representation—a practical approach," in Proc. of *First International Workshop on Image Databases and Multi Media Search,* 1996.
[24] W. Niblack et al., "The QBIC project: Querying images by content using color, texture and shape," in Proc.*SPIE*, vol. 1908, San Jose, CA, pp.173-187, Feb. 1993.