Content based image retrieval techniques – Issues, analysis and the state of the art

Darshak G. Thakore[1], A. I. Trivedi[2]

[1]*B V M Engineering College, Vallabh Vidyanagar, Gujarat*

[2]IEEE Member, *M. S. University, Vadodara,* Gujarat

Abstract

The revolutionary internet and digital technologies have imposed a need to have a system to organize abundantly available digital images for easy categorization and retrieval. The need to have a versatile and general purpose content based image retrieval (CBIR) system for a very large image database has attracted focus of many researchers of information-technology-giants and leading academic institutions for development of CBIR techniques. These techniques encompass diversified areas, viz. image segmentation, image feature extraction, representation, mapping of features to semantics, storage and indexing, image similarity-distance measurement and retrieval - making CBIR system development a challenging task. The paper addresses and analyses challenges & issues of CBIR techniques/systems, evolved during recent years covering various methods for segmentation; edge, boundary, region, color, texture, and shape based feature extraction; object detection and identification. The state of the art techniques are reviewed and future scope is cited.

Key Words: CBIR, Image feature extraction, Image analysis, Image retrieval, Image search, Image similarity

1. Introduction

Retrieval of *required-query-similar* images from abundantly available / accessible digital images is a challenging need of today. The image retrieval techniques based on visual image content has been in-focus for more than a decade. Many web-search-engines retrieve similar images by searching and matching textual metadata associated with digital images. For better precision of the retrieved resultant images, this type of search requires associating meaningful image-descriptive-text-labels as metadata with all images of the database. Manual image labeling, known as manual image annotation, is practically difficult for exponentially increasing image database. The image search results, appearing on the first page for fired text query *rose black*, are shown in Figure 1 for leading web search engines Google, Yahoo and AltaVista.

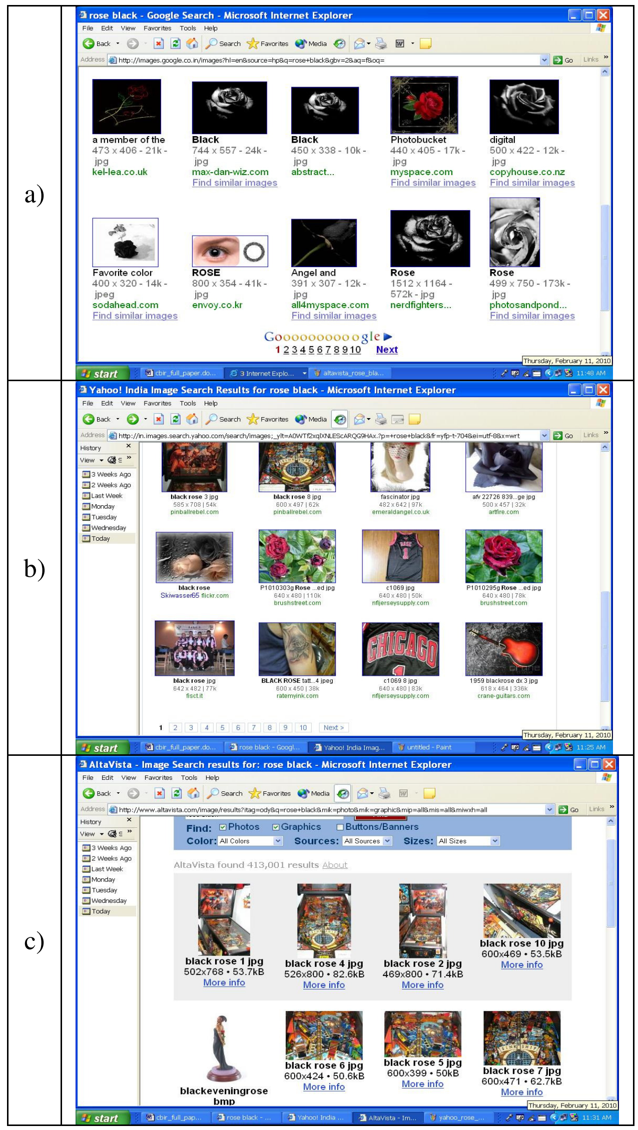

Figure 1. Image search results for query – rose black a) Google b) Yahoo c) AltaVista

Many resultant images of Figure 1 lack semantic matching with the query, showing vast scope of research leading to improvements in the state-of-art-techniques. The need evolved two solutions – automatic image annotation and content based image retrieval. The content based image retrieval techniques aim to respond to a query image (or sketch) with query-similar resultant images obtained from the image database. The database images are preprocessed for extracting and then storing – indexing corresponding image features. The query image also gets processed for extracting features which are compared with features of database images by applying appropriate similarity measures for retrieving query-similar-images.

2. Issues

The biggest issue for CBIR system is to incorporate versatile techniques so as to process images of diversified characteristics and categories. Many techniques for processing of low level cues are distinguished by the characteristics of domain-images. The performance of these techniques is challenged by various factors like image resolution, intra-image illumination variations, non-homogeneity of intra-region and inter-region textures, multiple and occluded objects etc. The other major difficulty, described as semantic-gap in the literature, is a gap between inferred understanding / semantics by pixel domain processing using low level cues and human perceptions of visual cues of given image. In other words, there exists a gap between mapping of extracted features and human perceived semantics. The dimensionality of the difficulty becomes adverse because of subjectivity in the visually perceived semantics, making image content description a subjective phenomenon of human perception, characterized by human psychology, emotions, and imaginations. The image retrieval system comprises of multiple inter-dependent tasks performed by various phases. Inter-tuning of all these phases of the retrieval system is inevitable for over all good results. The diversity in the images and semantic-gap generally enforce parameter tuning & threshold-value specification suiting to the requirements. For development of a real time CBIR system, feature processing time and query response time should be optimized. A better performance can be achieved if feature-dimensionality and space complexity of the algorithms are optimized. Specific issues, pertaining to application domains are to be addressed for meeting application-specific requirements. Choice of techniques, parameters and threshold-values are many a times application domain specific e.g. a set of techniques and parameters producing good results on an image database of natural images may not produce equally good results for medical or microbiological images.

3. Image features

Various techniques for extraction and representation of image features like histograms – local (corresponding to regions or sub-image) or global , color layouts, gradients, edges, contours, boundaries & regions, textures and shapes have been reported in the literature.

Histogram is one of the simplest image features. Despite being invariant to translation and rotation about viewing axis, lack of inclusion of spatial information is its major draw back. Many totally dissimilar images may have similar histograms as spatial information of pixels is not reflected in the histograms. Consequently, many histogram refinement techniques have been reported in the literature. Histogram intersection based method for comparing model and image histograms was proposed in [1] for object identification. Histogram refinement based on color coherence vectors was proposed in [3]. The technique considers spatial information and classifies pixels of histogram buckets as coherent if they belong to a small region and incoherent otherwise. Though being computationally expensive, the technique improves performance of histogram based matching. Color correlogram feature for images was proposed in [2] which take into account local color spatial correlation as well as global distribution of this spatial correlation. The correlogram gives the change of spatial correlation of pairs of colors with distance and hence performs well over classical histogram based techniques. A modified histogram based technique to incorporate spatial layout information of each color with annular, angular and hybrid histograms has been proposed in [4]. In [5], cumulative histogram and respective distances for image similarity measures, overcoming quantization problem of the histogram bins was proposed. The representation of color distribution features for each color channel based on average, variance and skewness, described as moments, for image similarity was also presented.

Various segmentation techniques based on edge detection, contour detection and region formation have been reported in the literature. These techniques, in general, process low level cues for deriving image features by following bottom-up approach. Automatic image segmentation is a very crucial phase as the overall performance of retrieval results significantly depends on the precision of the segmentation. The most difficult task for any automatic image segmentation algorithm is to avoid under and over segmentation of images, possessing diversified characteristics. Hence, for required scale of segmentation, parameter tuning or threshold adjustment becomes unavoidable for versatile image segmentation algorithms.

Directional changes in color and texture have been identified in [10], using predictive color model to detect boundaries by iteratively propagating *edge flow*. This iterative method is computationally expensive because of processing of low level cues at all pixels for given scale. A novel hierarchical classification frame work based approach for boundary extraction with Ulrtametric

Contour Maps UCM - representing geometric structure of an image has been proposed in [7]. A generic grouping algorithm based on Oriented Watershed Transform and UCM [7] has been proposed in [6] to form a hierarchical region tree, finally leading to segmentation. The method enforces bounding contour closures, avoiding leaks - a root cause of under segmentation. Exhaustive precision-recall evaluation of OWT-UCM technique for different scales also has been presented. Region based image retrieval, incorporating graphs, multiple low level labels and their propagation, multilevel semantic representation and support vector machine has been proposed in [14], implying effectiveness of the method.

Various techniques based on generalized Hough transform and Fourier descriptors have been reported in the literature for shape and object boundary detection. A review of methods for shape comparison has been reported in [11]. Active contour model – snake has been used in [12] for interactive interpretation, where user-imposed constraint forces guide the snake to feature of interest. Many variations based on active contour methods have been found in literature. The boundary detection precision of active contour based methods is generally sensitive to seed-points or seed-contours, if not provided properly, snakes may not converge to true object boundaries. Recursive mean shift procedure based analysis of multimodal feature space and delineation of arbitrarily shaped cluster can be found in [9]. Scale invariant local shape features with chains of k-connected roughly straight family of contour segments has been used for object class detection in [13].

Many relevance feed back techniques have been proposed in literature to bridge the semantic gap by specifying positive and negative feed backs given by the user for refinement of results. A relevance feedback based interactive image retrieval approach to address issues of semantic-gap and subjectivity of human perception of visual contents was introduced in [23], which showed significant improvement in the results. In [40], orthogonal complement component based relevance feed back technique is proposed that does not treat positive and negative feed backs equivalently, as the former share homogenous concepts where as latter do not. Generalized Bayesian learning framework with target query and a user conception based user-model has been proposed in [22] where target distribution, target query and matching criteria have been updated at every feed back step.

A fuzzy approach based CBIR, named FIRST - Fuzzy Image Retrieval SysTem, has been proposed in [15] to handle the vagueness in the user queries and inherent uncertainty in image representation, similarity measure, and relevance feedback incorporating fuzzy attributed relational graph comparisons for similarity measures. Contour and texture cues have been exploited simultaneously in [8] using intervening contour frame work and textons for image segmentation with spectral graph theoretic framework of normalized cuts. Perceptual grouping of block based visual patterns using modified Hough transform for object search technique in heterogeneous cluster-oriented CBIR with load balancing implementation has been reported in [21]. Two new texture features - Block difference of inverse probabilities (BDIP), measuring local brightness variations & block variation of local correlation coefficients (BVLC), and measuring local texture smoothness have been used in [16] and the combination of BDIP and BVLC moments for image retrieval improves performance compared to wavelet moments. Evolutionary group algorithm to optimize the quantization thresholds of the wavelet-correlogram has been reported in [17].

4. CBIR systems

A brief summary of some of the CBIR systems has been presented in this section. QBIC - Query By Image Content system, developed by IBM, makes visual content similarity comparisons of images based on properties such as color percentages, color layout, and textures occurring in the images. The query can either be example images, user-constructed sketches and drawings or selected color and texture patterns [26] [27]. The IBM developed QBIC technology based Ultimedia Manager Product for retrieval of visually similar images [28]. Virage [35] and Excalibur are other developers of commercial CBIR systems.

VisualSEEk **-** a joint spatial-feature image search engine developed at Columbia university performs image similarity comparison by matching salient color regions for their colors, sizes and absolute & relative spatial locations[29][30]. Photobook developed at Media Laboratory, Massachusetts Institute of Technology – MIT for image retrieval based on image contents where in color, shape and texture features are matched for euclidean, mahalanobis, divergence, vector space angle, histogram, Fourier peak, and wavelet tree distances. The incorporation of interactive learning agent, named *FourEyes* for selecting & combining feature-based models has been a unique feature of Photobook [31]. MARS - Multimedia Analysis and Retrieval Systems [32] and FIRE- Flexible Image Retrieval Engine [33] incorporate relevance feed back from the user for subsequent result refinements. Similar images are retrieved based on color features, Gabor filter bank based texture features, Fourier descriptor based shape features and spatial location information of segmented image regions in NeTra [34]. For efficient indexing, color features of image regions has been represented as subsets of color code book containing total of 256 colors. The frame work proposed in [10] has been incorporated for image segmentation in NeTra.

PicSOM (Picture & Self-organizing Map) was implemented using tree structured SOM, where SOM was used for image similarity scoring method [36]. Visual content descriptors of MPEG-7 (Moving Pictures Expert Group Multimedia Content Description Interface) were used in PicSOM [36] for CBIR techniques and performance comparison with Vector Quantization based system was proposed in [37]. Incorporation of relevance feedback in it caused improvements in the precision of results of Picsom. SIMPLIcity – Semantics-sensitive Integrated Matching for Picture Libraries incorporates integrated region matching methodology for overcoming issues related to improper image segmentation. The segmented images are represented as sets of regions. These regions, roughly corresponding to objects are characterized by their colors, shapes, textures and locations. The image search is narrowed-down by applying image-semantic-sensitive categorization for better retrieval performance [38]. The online demo of SIMPLIcity is available at [39]. Comparison of the mean average precision of three content based image retrieval methodologies have been presented in [18], indicating improvements in the performance over last few years. Performance comparison of query by visual example and query by semantic example has been reported in [18], demonstrating superior performance of the latter. As reported, the content based image retrieval methodologies have evolved from modeling visual appearance, to learning semantic models and finally to making inferences using semantic spaces. Performance comparison of minimum probability of error retrieval frame work based query by visual example and query by semantic example has been reported in [20], concluding *semantic representations of images have an intrinsic benefit for image retrieval*. Elaborative study of query by semantic example addressing structure of semantic space and effect of low level visual features & high level semantic features on over all performance of CBIR system has been reported in [19].

A comparative survey of CBIR systems / techniques have been reported in [24][25].

5. Discussion

The road map of development of CBIR techniques began with simple primitive features based indexing methodologies that later got enhanced with combinational features. Two major issues, semantic-gap and subjectivity of semantics are addressed by the state of the art techniques. Many state of the art techniques incorporate iterative relevance feed back from user for refinement of results. Semantic gap bridging approaches based on fuzzy, evolutionary and neural network have also been reported. Hierarchical approaches for feature extraction and representations achieve hierarchical abstraction; help matching semantics of visual perception of human beings. Several modern techniques focus on improvements on processing of low level cues so as to precisely extract features. We intend to propose that prominent boundary detection based hierarchical approach with region feature extraction would significantly improve the quality of retrieval results. Many state of the art techniques suggest that semantic domain based image retrieval systems, comparing *meaningful* concepts improve quality of retrieved image set. Effective learning and inferring of meaningful concepts may get proved critical for such systems.

The state of the art image retrieval techniques have a scope of under-going significant technical evolution.

6. References

[1] M. Swain and D. Ballard, "Color indexing", *International* Journal *of Computer Vision*, 7(1), 1991, pp. 11–32.

[2] J. Huang, S. R. Kumar, M. Mitra, W. Zhu and R. Zabih, "Image Indexing Using Color Correlograms", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 1997, pp. 762 – 768.

[3] G. Pass and R. Zabih, "Histogram Refinement for Content Based Image Retrieval", *3rd IEEE Workshop on Applications of Computer Vision, WACV* , 1996 , pp. 96-102.

[4] A. Rao, R. K. Srihari and Z. Zhang, "Spatial Color Histograms for Content-Based Image Retrieval", *11th IEEE International Conference on Tools with Artificial Intelligence,* 1999, pp. 183 – 186.

[5] M. A. Stricker and M. Orengo, "Similarity of color images", Proc. *of the SPIE conference on the Storage and Retrieval for Image and Video Databases III*, 1995, pp. 381–392.

[6] P. Arbel´aez, M. Maire, C. Fowlkes, and J. Malik, "From Contours to Regions: An Empirical Evaluation", *CVPR* 2009, pp. 2294-2301.

[7] P. Arbel´aez, "Boundary Extraction in Natural Images Using Ultrametric Contour Maps", *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06*), 2006, pp. 182-182.

[8] J. Malik, S Belongie, T Leung And J Shi, "Contour and Texture Analysis for Image Segmentation", *International Journal of Computer Vision*, Vol. 43, Issue 1, 2001, pp. 7-27.

[9] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis", *IEEE transaction on PAM*I, Vol. 24, No 5, May 2002, pp. 603-619.

[10] W. Ma and B. S. Manjunath, "EdgeFlow: A Technique for Boundary Detection and Image Segmentation", *IEEE transaction on, Image Processing*, Vol. 9, Issue 8, August 2000, pp. 1375-1388.

[11] R. C. Veltkamp and M. Hagendoorn, "State-of-the-Art in Shape Matching", *Multimedia Search: State of the Art*, Springer-Verlag, 2000.

[12] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active Contour Models", *IJCV,* 1988, pp. 321-331.

[13] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of Adjacent Contour Segments for Object Detection", *IEEE transaction on PAMI*, Volume 30 , Issue 1, Jan. 2008, pp. 36-51.

[14] F. Li, Q. Dai, W. Xu and G. Er, "Multilabel Neighborhood Propagation for Region-Based Image Retrieval", *IEEE Transactions On Multimedia*, Vol. 10, No. 8, December 2008, pp. 1592-1604.

[15] R. Krishnapuram, S. Medasani, S. Jung, Y. Choi, and R. Balasubramaniam, "Content-Based Image Retrieval Based on a Fuzzy Approach", *IEEE Transactions On Knowledge And Data Engineering*, Vol. 16, No. 10, October 2004, pp. 1185-1199.

[16] Y. D. Chun, S. Y. Seo, and N. C. Kim, "Image Retrieval Using BDIP and BVLC Moments", *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 13, No. 9, September 2003, pp. 951-957.

[17] M. Saadatmand-Tarzjan and H. A. Moghaddam, "A Novel Evolutionary Approach for Optimizing Content-Based Image Indexing Algorithms", *IEEE Transactions On Systems, Man, And Cybernetics—Part B: Cybernetics*, Vol. 37, No. 1, February 2007, pp. 139-153.

[18] N. Vasconcelos, "From Pixels to Semantic Spaces: Advances in Content-Based Image Retrieval", *Computer* Volume: 40, Issue: 7, 2007, pp. 20-26.

[19] N. Rasiwasia and N. Vasconcelos, "A Study of Query by Semantic Example", *3rd International Workshop on Semantic Learning and Applications in Multimedia*, Anchorage, June 2008, pp. 1-8.

[20] N. Rasiwasia, P. J. Moreno and N. Vasconcelos, "Bridging the Gap: Query by Semantic Example", *IEEE Transactions On Multimedia*, Vol. 9, No. 5, August 2007, pp. 923-938.

[21] S. Cheng, W. Huang, Y. Liao and D. Wu, "A Parallel CBIR Implementation Using Perceptual Grouping Of Block-based Visual Patterns", *IEEE International Conference on Image Processing – ICIP*, 2007, pp. V - 161 - V - 164.

[22] C. Hsu and C. Li, "Relevance Feedback Using Generalized Bayesian Framework With Region-Based Optimization Learning", *IEEE Transactions On Image Processing*, Vol. 14, No. 10, October 2005, pp. 1617- 1631.

[23] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval", *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 8, No. 5, September 1998, pp. 644 – 655.

[24] R. C. Veltkamp and M. Tanase, "Content-Based Image Retrieval Systems: A Survey", http://aa-lab.cs.uu.nl/cbirsurvey/cbir-survey/

[25] A. W. Smeulders, M. M. Worring, A. Gupta and R. Jain, "Content-Based Image Retrieval at the End of the Early Years", *IEEE Transaction on Pattern Anal. Machine Intelligence,* Vol. 22 No.12, 2000, pp. 1349-1380.

[26] The QBIC web site: http://wwwqbic.almaden.ibm.com/

[27] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafher, D. Lee, D. Petkovie, D. Steele, and P. Yanker, "Query by image and video content: the QBIC system", *Computer*, Volume: 28, Issue: 9, 1995, pp. 23-32

[28] R. Barber, M. Flickner, J. Hafner, D. Lee, W. Niblack, D. Petkovic, J. Ashley, T. McConnell, J. Ho, J. Jang, D. Berkowitz, P. Yanker, M. Vo, D. Ilaas, D. Lassig, S. Tate, A. Chang, P. van Houten, J. Chang, T. Petersen, D. Lutrell, M. Snedden, P. Faust, C. Matteucci, M. Rayner, R. Peters, W. Beck, J. Witsett, **"**Ultimedia Manager: Query By Image Content And Its Applications", *Digest of Papers, Compcon Spring '94*, 1994, pp. 424-429.

[29] J. R. Smith and S. F. Chang, "VisualSEEk: A Fully Automated Content-Based Image Query System", *Proc. ACM Int. Conf. Multimedia*, Boston, MA, 1996, pp. 87–98.

[30] The VisualSEEk web site: http://www.ee.columbia.edu/ln/dvmm/researchProjects/MultimediaIndexing/VisualSEEk/VisualSEEk.htm

[31] The Photobook web site: http://vismod.media.mit.edu/vismod/demos/photobook/ph6

[32] The MARS web site: http://www.ifp.illinois.edu/~qitian/MARS.html

[33] The fire web site: http://www-i6.informatik.rwth-aachen.de/~deselaers/cgi_bin/fire.cgi

[34] W.Y. Ma and B.S. Manjunath, "NeTra: A Toolbox For Navigating Large Image Databases", *IEEE Int. Conf. on Image Processing*, 1997, pp. 568-571.

[35] The Virage web site: http://www.virage.com

[36] J. Laaksonen, M. Koskela and E. Oja, "PicSOM: Self-Organizing Maps For Content-Based Image Retrieval", *International Joint Conference on Neural Networks, IJCNN '99,* Volume 4, July 1999, pp. 2470 – 2473.

[37] J. Laaksonen, M. Koskela and E. Oja, "PicSOM Self Organizing Image Retrieval With MPEG-7 Content Descriptors", *IEEE Transactions On Neural Networks*, Vol. 13, No. 4, July 2002, pp. 841-853.

[38] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries", *IEEE transaction on PAMI*, Volume 23, No 9, Sept 2001, pp. 947-963.

[39] The SIMPLIcity web site: http://wang14.ist.psu.edu/cgi-bin/zwang/regionsearch_show.cgi

[40] D. Tao, X. Tang, and X. Li "Which Components are Important for Interactive Image Searching?", *IEEE Transactions On Circuits And Systems For Video Technology*, Vol. 18, No. 1, January 2008, pp. 3-11.