State of Art Literature Survey on Content base Image Retrieval by Multi Features

Arpita Mathur
Department of Computer Science, Lachoo Memorial College of Sc. & Tech.
A-Sector, Shastri Nagar, Jodhpur (India)

Rajeev Mathur
Department of Computer Science, Lachoo Memorial College of Sc. & Tech.
A-Sector, Shastri Nagar, Jodhpur (India)

ABSTRACT
Rapid growth of World Wide Web has increased the interest towards image retrieval. Different groups need to find a desired image from a collection. The users may require access to the images, based on primitive features, such as color, texture or shape, or associated text. The technology to access these images has also accelerated phenomenally. The current approaches are broad and inter-disciplinary, mainly focused on three aspects of image research which are text-based retrieval, content-based retrieval and interactive based image retrieval. Recently, Content-Based Image Retrieval (CBIR) has become an active research area. This paper gives the literature survey for CBIR which explains rapid growth in this field. It briefly discusses the work done by different researchers.

General Terms
Pattern Recognition, Algorithms, Machine Learning, Image Processing, Image Retrieval, Content Based Image Retrieval, Multi Features. text based retrieval, interactive based image retrieval.

Keywords
Pattern Recognition, Algorithms, Machine Learning, Image Processing, Image Retrieval, Content Based Image Retrieval, Features, Texture, Shape, Entropy, text based retrieval, interactive based image retrieval.

CONTRIBUTIONS SO FAR ON CBIR
The earliest work on Content Based Image Retrieval was done by Ning-San Chang and
King-Sun Fu[1] in their paper Query-by-Pictorial-Example. They introduced Query-by-Pictorial-Example as a relational query language for manipulating queries regarding pictorial relations as well as conventional relations. In addition to the manipulating capabilities of the conventional query languages, queries could also be expressed in terms of pictorial examples through a display terminal. Example queries were used to illustrate the language facilities.

T. Joseph, A.F. Cardenas [4] presented a corresponding high-level query language, PICQUERY, illustrated through examples. PICQUERY has been designed with a flavor similar to QBE as the highly nonprocedural and conservational language for the pictorial database management system PICDMS. PICQUERY and a relational QBE-like language formed the language by which a user could access conventional relational databases and at the same time pictorial databases managed by PICDMS or other robust PDBMS. This language interface is part of an architecture aimed toward data heterogeneity transparency over pictorial and nonpictorial databases.

Eden and Unse [5] described an approach for unsupervised segmentation of textured images. Local texture properties are extracted using local linear transforms that have been optimized for maximal texture discrimination. Local statistics (texture energy measures) are estimated at the output of an equivalent filter bank by means of a nonlinear transformation (absolute value) followed by an iterative Gaussian smoothing algorithm. This procedure generated a multiresolution sequence of feature planes with a half-octave scale progression. A feature reduction technique was then applied to the data and was determined by simultaneously diagonalizing scatter matrices evaluated at two different spatial resolutions. This approach provided a good approximation of R.A. Fisher's (1950) multiple linear discriminants and has the advantage of requiring no a priori knowledge. This feature reduction methods was an improvement on the commonly used Karhunen-Loeve transform and allows efficient texture segmentation based on simple thresholding.

John R. Smith and Shih-Fu Chang [7] implemented a project "Content-based" image /video retrieval system" for the World Wide Web. The application, called VisualSEEk, provides a suite of tools with which a person may search for and retrieve images and videos over the Web. The person formulates visual queries by using the VisualSEEk interface tools to illustrate salient visual properties of the images desired. They query is sent to the server which finds and retrieves to the user the images that best match the visual description in the query. For the scope of the initial project the usable image features include image color and the spatial layout of color regions. For example, the features are used as follows: to retrieve images of "sunsets" from the image archive, one possible query can be constructed by sketching a yellow circle near the center of the image (for the sun) and filling the upper part of the image with orange (for the sky). The images that best match this query contained colored regions that closely match.

Anil Jain and Lin Hong [9] have developed a prototype biometrics system which integrates faces and fingerprints. The system overcomes the limitations of face recognition systems as well as fingerprint verification systems. The integrated prototype system operates in the identification mode with an admissible response time. The identity established by the system is more reliable than the identity established by a face recognition system. In addition, the proposed decision fusion scheme enables performance improvement by integrating multiple cues with different confidence measures. Experimental results demonstrate that their system performs very well. It meets the response time as well as the accuracy requirements.
Using binary Bayesian classifiers, Figueiredo et al [10] attempted to capture high-level concepts from low-level image features under the constraint that the test image does belong to one of the classes. Specifically, they consider the hierarchical classification of vacation images; at the highest level, images

are classified as indoor or outdoor; outdoor images are further classified as city or landscape; finally, a subset of landscape images was classified into sunset, forest, and mountain classes. They demonstrated that a small vector quantizer (whose optimal size is selected using a modified MDL criterion) can be used to model the class-conditional densities of the features, required by the Bayesian methodology. The classifiers have been designed and evaluated on a database of 6931 vacation photographs. Their system achieved a classification accuracy of 90.5% for indoor/outdoor, 95.3% for city/landscape, 96.6% for sunset/forest and mountain, and 96% for forest/mountain classification problems. They further developed a learning method to incrementally train the classifiers as additional data become available. They also show preliminary results for feature reduction using clustering techniques.

Grimson et al [12] in their paper "Spatial template extraction for image retrieval by region matching" presented a template and its relation extraction and estimation (TREE) algorithm for indexing images from picture libraries with more semantics-sensitive meanings. This algorithm can learn the commonality of visual concepts from multiple images to give a middle-level understanding about image contents. In this approach, each image is represented by a set of templates and their spatial relations as keys to capture the essence of this image. Each template is characterized by a set of dominant regions, which reflect different appearances of an object at different conditions and can be obtained by the template extraction and analysis (TEA) algorithm through region matching. The spatial template relation extraction and measurement (STREAM) algorithm was then proposed for obtaining the spatial relations between these templates. Due to the nature of a template, which can represent object's appearances at different conditions, the proposed approach owns better capabilities and flexibilities to capture image contents than traditional region-based methods. In addition, through maintaining the spatial layout of images, the semantic meanings of the query images can be extracted and lead to significant improvements in the accuracy of image retrieval. Since no time-consuming optimization process is involved, the proposed method learns the visual concepts extremely fast. Experimental results provided to prove the superiority of the proposed method.

Valova et al [13] presented their work in the area of image organization and retrieval in Image Databases using global color features and spatial color distribution of images. They suggested extending the use of image histograms to characterize the global and local color properties of an image and to preserve its intrinsic geometric information. They have introduced an algorithm for image organization and retrieval by color features, presented as ORCF. The proposed algorithm is robust in the sense that it can deal with scale and rotation variances in images. On the basis of this algorithm they have developed a prototype system for content base image organization and retrieval.

Ghebreab et at [14] proposed a method for concept-based medical image retrieval that is a superset of existing semantic-based image retrieval methods. They conceive of a concept as an incremental and interactive formalization of the user's conception of an object in an image. The premise is that such a concept is closely related to a user's specific preferences and subjectivity and, thus, allows to deal with the complexity and content-dependency of medical image content. They describe an object in terms of multiple continuous boundary features and represent an object concept by the stochastic characteristics of an object population. A population-based incrementally learning technique, in combination with relevance feedback, is then used for concept customization. The user determines the speed and direction of concept customization using a single parameter that defines the degree of exploration and exploitation of the search space. Images are retrieved from a database in a limited number of steps based upon the customized concept. To demonstrate their method they have performed concept-based image retrieval on a database of 292 digitized X-ray images of cervical vertebrae with a variety of abnormalities. The results show that their method produces precise and accurate results when doing a direct search. In an open-ended search their method efficiently and effectively explores the search space.

In paper titled "A Dynamic User Concept Pattern Learning Framework for Content-Based Image Retrieval " the authours Zhang et al [16] says it is crucial to effectively discover users' concept patterns through an acquired understanding of the subjective role played by humans in the retrieval process for such systems. A learning and retrieval framework is used to achieve this. It seamlessly incorporates multiple instance learning for relevant feedback to discover users concept patterns-especially in the region of greatest user interest. It also maps the local feature vector of that region to the high-level concept pattern. This underlying mapping can be progressively discovered through feedback and learning. The user guides the retrieval systems learning process using his/her focus of attention. Retrieval performance is tested to establish the feasibility and effectiveness of the proposed learning and retrieval framework.

In the paper "Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples" the authours Kherfi et al [17] presented a new RF framework based on a feature selection algorithm that nicely combines the advantages of a probabilistic formulation with those of using both the positive example (PE) and the negative example (NE). Through interaction with the user their algorithm learns the importance they assigns to image features, and then applied the results obtained to define similarity measures that correspond better to his judgement. The use of the NE allows images undesired by the user to be discarded, thereby improving retrieval accuracy. As for the probabilistic formulation of the problem, it presents a multitude of advantages and opens the door to more modeling possibilities that achieve a good feature selection. It makes it possible to cluster the query data into classes, choose the probability law that best models each class, model missing data, and support queries with multiple PE and/or NE classes. The basic principle of their algorithm is to assign more importance to features with a high likelihood and those which distinguish well between PE classes and NE classes. The proposed algorithm was validated separately and in image retrieval context, and the experiments show that it performs a good feature selection and contributes to improving retrieval effectiveness.

Basak et al [18] designed a content-based image retrieval system where multiple query examples can be used to indicate the need to retrieve not only images similar to the individual examples, but also those images which actually represent a combination of the content of query images. They proposed a scheme for representing content of an image as a combination of features from multiple examples. This scheme is exploited for developing a multiple example-based retrieval engine. They have explored the use of machine learning techniques for generating the most appropriate feature combination scheme for a given class of images. The combination scheme can be used for developing purposive query engines for specialized image databases. Here, they have considered facial image databases. The effectiveness of the image retrieval system is experimentally demonstrated on different databases.

To investigate strategies to alleviate the vocabulary problem, Chai et al [19] examined the role of user term feedback in targeted image search that is based on text-based image retrieval. Term feedback refers to the feedback from a user on specific terms regarding their relevance to a target image. Previous studies have indicated the effectiveness of term feedback in interactive text retrieval. In their experiments on text-based image retrieval, the term feedback has not been shown to be effective. The results indicate that, although term feedback has a positive effect by allowing users to identify more relevant terms, it also has a strong negative effect by providing more opportunities for users to specify irrelevant terms. To understand these different effects and their implications, the article further analyzes important factors that contribute to the utility of term feedback and discusses the outlook of term feedback in interactive text-based image retrieval.

In the paper titled "Multilabel Neighborhood Propagation for Region-Based Image Retrieval" the authours Fei et al [20] designed a framework based on multilabel neighborhood propagation is proposed for RBIR, which can be characterized by three key properties: (1) For graph construction, in order to determine the edge weights robustly and automatically, mixture distribution is introduced into the Earth mover's distance (EMD) and a linear programming framework is involved. (2) Multiple low-level labels for each image can be obtained based on a generative model, and the correlations among different labels are explored when the labels are propagated simultaneously on the weighted graph. (3) By introducing multilayer semantic representation (MSR) and support vector machine (SVM) into the long-term learning, more exact weighted graph for label propagation and more meaningful high-level labels to describe the images can be calculated. Experimental results, including comparisons with the state-of-the-art retrieval systems, demonstrate the effectiveness of their proposal.

Cholleti et al [21] defined localized content-based image retrieval as a CBIR task where the user is only interested in a portion of the image, and the rest of the image is irrelevant. In their paper they present a localized CBIR system, ACCIO, that uses labeled images in conjunction with a multiple-instance learning algorithm to first identify the desired object and weight the features accordingly, and then to rank images in the database using a similarity measure that is based upon only the relevant portions of the image. A challenge for localized CBIR is how to represent the image to capture the content. They present and compare two novel image representations, which extend traditional segmentation-based and salient point-based techniques respectively, to capture content in a localized CBIR setting.

According to Lei et at [22] conventional content-based image retrieval (CBIR) schemes employing relevance feedback may suffer from some problems in the practical applications. First, most ordinary users would like to complete their search in a single interaction especially on the Web. Second, it is time consuming and difficult to label a lot of negative examples with sufficient variety. Third, ordinary users may introduce some noisy examples into the query. This correspondence explores solutions to a new issue that image retrieval using unclean positive examples. In the proposed scheme, multiple feature distances are combined to obtain image similarity using classification technology. To handle the noisy positive examples, a new two-step strategy is proposed by incorporating the methods of data cleaning and noise tolerant classifier. The extensive experiments carried out on two different real image collections validate the effectiveness of the proposed scheme.

A shape-based, hierarchical part-template matching approach to simultaneous human detection and segmentation combining local part-based and global shape-template-based schemes was proposed by Davis et al [24].The approach relies on the key idea of matching a part-template tree to images hierarchically to detect humans and estimate their poses. For learning a generic human detector, a pose-adaptive feature computation scheme is developed based on a tree matching approach. Instead of traditional concatenation-style image location-based feature encoding, we extract features adaptively in the context of human poses and train a kernel-SVM classifier to separate human/nonhuman patterns. Specifically, the features are collected in the local context of poses by tracing around the estimated shape boundaries. They also introduce an approach to multiple occluded human detection and segmentation based on an iterative occlusion compensation scheme. The output of their learned generic human detector can be used as an initial set of human hypotheses for the iterative optimization. They evaluate their approaches on three public pedestrian data sets (INRIA, MIT-CBCL, and USC-B) and two crowded sequences from Caviar Benchmark and Munich Airport data sets.

A perception-based approach to content-based image representation and retrieval is proposed by Abbadeni [26]. They consider textured images and propose to model their textural content by a set of features having a perceptual meaning and their application to content-based image retrieval. They present a new method to estimate a set of perceptual textural features, namely coarseness, directionality, contrast, and busyness. The proposed computational measures can be based upon two representations: the original images representation and the autocorrelation function (associated with original images) representation. The set of computational measures proposed is applied to content-based image retrieval on a large image data set, the well-known Brodatz database. Experimental results and benchmarking show interesting performance of their approach. First, the correspondence of the proposed computational measures to human judgments is shown using a psychometric method based upon the Spearman rank-correlation coefficient. Second, the application of the proposed computational measures in texture retrieval shows interesting results, especially when using results fusion returned by each of the two representations. Comparison is also given with related works and show excellent performance of their approach compared to related approaches on both sides: correspondence of the proposed computational measures with human judgments as well as the retrieval effectiveness.

Decheng et al [27] aswered the questions like how do we retrieve cartoon characters accurately or how to synthesize new cartoon clips smoothly and efficiently from the cartoon library? In this paper, they consider multiple features from different views, i.e., color histogram, Hausdorff edge feature, and skeleton feature, to represent cartoon characters with different colors, shapes, and gestures. Each visual feature reflects a unique characteristic of a cartoon character, and they are complementary to each other for retrieval and synthesis. Secondly, to combine the three visual features by simply concatenating them into a long vector, it will end up with the so-called â€œcurse of dimensionality, let alone their heterogeneity embedded in different visual feature spaces. Here, they introduce a semisupervised multiview subspace learning (semi-MSL) algorithm, to encode different features in a unified space. Specifically, under the patch alignment framework, semi-MSL

uses the discriminative information from labeled cartoon characters in the construction of local patches where the manifold structure revealed by unlabeled cartoon characters is utilized to capture the geometric distribution. The experimental evaluations based on both cartoon character retrieval and clip synthesis demonstrate the effectiveness of the proposed method for cartoon application. Moreover, additional results of content-based image retrieval on benchmark data suggest the generality of semi-MSL for other applications.

According to Wen et al [28] it is important to use visual information in order to solve the ambiguity in text-based image retrieval. In there paper they proposed a novel Internet image search approach. It only requires the user to click on one query image with minimum effort and images from a pool retrieved by text-based search are reranked based on both visual and textual content. Thier key contribution is to capture the users' search intention from this one-click query image in four steps. 1) The query image is categorized into one of the predefined adaptive weight categories which reflect users' search intention at a coarse level. Inside each category, a specific weight schema is used to combine visual features adaptive to this kind of image to better rerank the text-based search result. 2) Based on the visual content of the query image selected by the user and through image clustering, query keywords are expanded to capture user intention. 3) Expanded keywords are used to enlarge the image pool to contain more relevant images. 4) Expanded keywords are also used to expand the query image to multiple positive visual examples from which new query specific visual and textual similarity metrics are learned to further improve content-based image reranking. All these steps are automatic, without extra effort from the user. This is critically important for any commercial web-based image search engine, where the user interface has to be extremely simple. Besides this key contribution, a set of visual features which are both effective and efficient in Internet image search are designed. Experimental evaluation shows that their approach significantly improves the precision of top-ranked images and also the user experince.

REFERENCES

[1] Chang, N.S., Fu K.S.,Query-by-Pictorial-Example, 1980, IEEE Transactions on Software Engineering.

[2] Herman, M.,Kanade, T.;Kuroe, Shigeru, "Incremental Acquisition of a Three-Dimensional Scene Model from Images", IEEE Transactions, May 1984,Volume : PAMI-6,Issue : 3, Pages : 331 to 340.

[3] Fu, K. S., "A Step Towards Unification of Syntactic and Statistical Pattern Recognition", IEEE Transactions, May 1986, Volume : PAMI-8 Issue : 3, Pages : 398 to 404

[4] Joseph, T., Cardenas, A. F., "PICQUERY: A High Level Query Language for Pictorial Database Management",May 1988 , vol. 14 no. 5, pp. 630-638

[5] Eden, M., Unser, M., "Multiresolution feature extraction and selection for texture segmentation", IEEE Transactions, Jul 1989, Volume : 11 Issue : 7, 717 to 728.

[6] Wu, C. M., Hsieh, K. S, Chen, Y. C., "Texture features for classification of ultrasonic liver images", IEEE Transactions, Jun 1992, Volume : 11 Issue : 2, 141 to 152.

[7] Smith, J. R., Chang, S. F., "VisualSEEk: A Content based image/video retrieval tool", 1997.

[8] Cass, T.A., "Robust affine structure matching for 3D object recognition", IEEE Transactions, Nov 1998, Volume : 20 Issue : 11, Pg 1265 to 1274

[9] Jain, A., Hong, L., "Integrating faces and fingerprints for personal identification", Dec 1998

[10] Figueiredo, M.A.T.,Hong-Jiang Zhang, Jain, A.K., Vailaya, A., "Image classification for content-based indexing ",IEEE Transactions, 2001.

[11] Sandeep, K., Rajagopalan, A. N., "Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information.", Proc. of Indian Conference on Computer Vision, Graphics and Image Processing, 2002.

[12] Grimson, W.E.L._2;Jun-Wei Hsieh_1, "Spatial template extraction for image retrieval by region matching", IEEE Transactions, Nov 2003.

[13] Valova, I., Rachevc, B., "Retrieval by Color Features in Image Databases.", Proc. of Adbis'04, University of Rousse, Budapest, Hungary, 2004.

[14] Ghebreab, S., Jaffe, C.C., Smeulders, A.W.M., "Population-based incremental interactive concept learning for image retrieval by stochastic string segmentations", IEEE Transactions , 2004

[15] Chen, Y., Bart, H. L., Teng, F., "A Content-Based Image Retrieval System for Fish Taxonomy", Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, 2005.

[16] Zhang C., C., Shyu M.L., Rubin, S.H., Chen, S. C., "User Concept Pattern Learning Framework for Content-Based Image Retrieval", IEEE Transactions, 2006.

[17] Kherfi, M.L., Ziou, D., "Relevance feedback for CBIR: a new approach based on probabilistic feature weighting with positive and negative examples", Apr 2006.

[18] Basak, J., Bhattacharya, K., Chaudhury, S., "Multiple Exemplar-Based Facial Image Retrieval Using Independent Component Analysis", IEEE Transactions, Dec 2006

[19] Chai J. Y., C. Zhang, R. Jin, "An Empirical Investigation of User Term Feedback in Textbased Targeted Image Search", ACM Transaction, Inform. Systems, Vol. 25, 2007, No 1.

[20] Er, G., Fei Li, Qionghai, Wenli Xu, W., "Multilabel Neighborhood Propagation for Region-Based Image Retrieval", IEEE Transactions, Dec 2008.

[21] Cholleti, S.R., Fritts, J.E., Goldman, S.A., Hui Zhang, Rahmani, R., " Localized Content-Based Image Retrieval", IEEE Transactions, Nov 2008

[22] Ye, L., Zang, J., "Content Based Image Retrieval Using Unclean Positive Examples", Image Processing, IEEE Transactions, Oct 2009.

[23] Camps-Valls, G.,Kanevski, M, Matasci, G., Tuia, D., "Document Title : Learning Relevant Image Features With Multiple-Kernel Classification", Geoscience and Remote Sensing, IEEE Transactions, Oct 2010

[24] Davis, L.S., Zhe Lin, " Document Title : Shape-Based Human Detection and Segmentation via Hierarchical Part-Template Matching", Pattern Analysis and Machine Intelligence, IEEE Transactions, Apr 2010.

[25] Harzallah, H., Jegou, H., Schmid, C., Verbeek, J., "Accurate Image Search Using the Contextual Dissimilarity Measure", IEEE Transactions, Jan. 2010.

[26] Abbadeni, N., "Computational Perceptual Features for Texture Representation and Retrieval", IEEE Transactions, Jan 2011.

[27] Dacheng Tao_3;Dongquan Liu_2;Hock Soon Seah_4;Jun Yu, " On Combining Multiple Features for Cartoon Character Retrieval and Clip Synthesis", IEEE Transactions, Oct 2012.

[28] Wen, F., Jingyu Cui, J., Ke Liu, K., Wang, X, Tang X., "IntentSearch: Capturing User Intention for One-Click Internet Image Search", IEEE Transactions , 2012.