2. CBIR Algorithms & CBIR Systems- Review & Analysis

2.1 Introduction

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 & indexing, image similarity-distance measurement and retrieval - making CBIR system development a challenging task.

The chapter addresses and analyses challenges & issues of CBIR techniques/systems, evolved during recent years covering various methods for segmentation; edge, boundary, region, color, texture & shape based feature extraction; object detection & identification and image retrieval. The chapter also covers overview of published surveys on CBIR techniques / system. An eminent contributor's point of view on CBIR, query response analysis of some existing CBIR systems, our observations on CBIR issues and finally our approaches for developed CBIR Algorithms and system are included in the chapter.

2.2 Research & Publication Trend Analysis

The publication trend for a period of 1995 – 2005 has been shown for 'image retrieval' in [Datta, 2008]. Extending the analysis up to year 2010 for 'image retrieval' article indexed by Google Scholar as on 17-01-2011 shows tremendous increase in the research interest and hence in the number of publications post year 2005, as shown in

Figure 5, left. The publication trend for last five years has been shown in Figure 5, right for 'image retrieval' as indexed by Google Scholar as on 17-01-2011. A slight fall in number of publications for year 2010 may be due to non-visit of Google-web crawler to publishing server till date i.e. early part of year 2011.

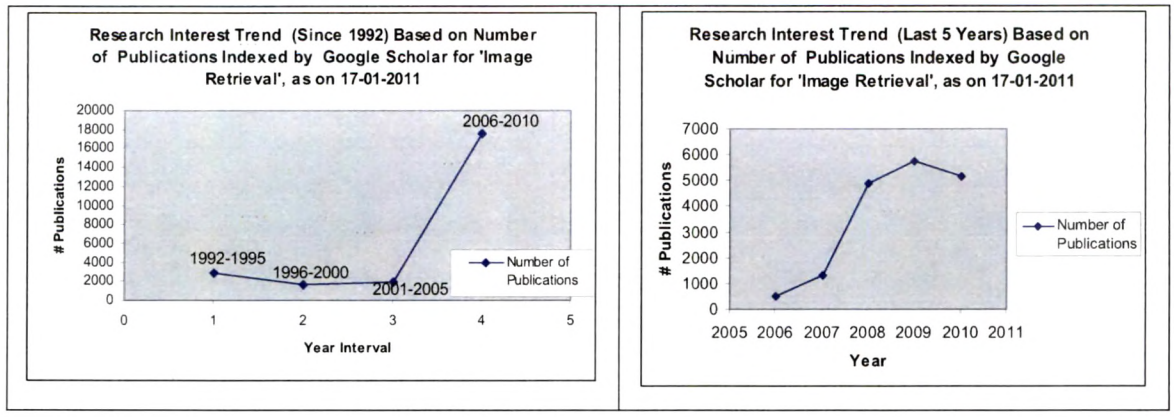

Figure 5. Research Interest Trend Based on Publications Indexed by Google Scholar for Image Retrieval. Left: Since 1992. Right: Last 5 Years.

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 diversities in the images and semantic-gaps 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.

2.3 Eminent Contributor's Point of View on CBIR

The issues presented in invited plenary talk at the *19th International Conference on Pattern Recognition*, held at Tampa, Florida during year 2008 by Theo Pavlidis - "Limitations of Content-based Image Retrieval" [Pavlidis, on line] are summarized below:

- Answer to the question to title of editorial in special issue of IEEE Proceedings [Hanjalic, 2008]: "*The Holy Grail of Multimedia Information Retrieval: So Close or Yet So Far Away?" was*
  - ***So close*** if published results are taken at face values
  - ***Yet so far away***, based on
    - closer look at the published results
    - the results of many test sites like GazoPa [GazoPa, on line] by Hitachi, TILTOMO [tiltomo, on line], ALIPR [aliper, on line] & many others
- The CBIR systems are designed to respond to broader queries than the specific.
- The methods and trainer / classifier are tested on too small number of images.
- Whether methods search for similar 2-D images or similar objects (3-D) of the image.
- Illumination / pose / viewpoint changes offer great challenges to simple feature based CBIR systems.
- The segmentation and salient point matching fail to deal pose / viewpoint variations of many objects.
- General similarity measures are needed for real-world CBIR.

- Computationally close images may have different perceptual meaning and computationally slightly differing images may have vast perceptual differences.
- "Accept that practically significant results for *real time general CBIR* cannot be obtained unless there are major breakthroughs both in image analysis and in computer architecture.
    - **As long as we do not have *general segmentation methods* that can identify objects on an image, it is unwise to pursue general CBIR."**

The current state of art and issues for automatic image annotation has also been presented by Theo Pavlidis [Pavlidis, on line] concluding direct human labeling is mandatory for proper image-tagging as automatic image tagging is very hard.

2.4 Query Response Analysis of Some of the Existing CBIR Systems

To illustrate the difficulties and complexity of CBIR, let us consider the query response of on-line demo of SIMPLIcity, developed at Penn State University, is available at http://wang14.ist.psu.edu/cgi-bin/zwang/regionsearch_show.cgi . The huge image database was consisting of 59895 images (as on 13-01-2011). The query of a deer image with ID 35600 gave no deer image in the response of displayed 32 images. Similarly, a horse-face, a close up image with id 35031 also resulted into query response of no horse image in displayed 32 images. The other on-line demo is available at http://amazon.ece.utexas.edu/~qasim/cires.htm having 50 image categories and around 30-50 images per category, developed at Computer and Vision Research centre, University of Texas, Austin. The image query for a wolf image named nat_anm_mam_wol_art_aw00001.jpg gave response with precision 3/20 (0.15). The enhanced version of CIRES provides browsing based on image-tags. The image browsing for tag provides 458 images (As on 14-01-2011) with precision 268/458 (0.58) which is available at http://cires.matthewriley.com/browse/view/7117. The precision in the enhanced version is better, but the result includes many images having human faces, perhaps due to significant weight to color in similarity matching. Given query response examples are only for the purpose of illustrating the challenges and difficulties involved in any CBIR algorithms / system and not for criticizing individual(s) or shortcomings of CBIR systems.

2.5 Overview of CBIR Surveys

The paragraph summarizes the points presented in the exhaustive survey by Datta et al. [Datta, 2008] carried out in 2008 on image retrieval covering about 300 theoretical and practical contributors. CBIR is cited in it as an emerging technology that in principal helps to organize digital pictures by their visual contents. The user perspective based on the clarity of user intention (browsing, surfing or searching), data scope (personal collection, domain specific, enterprise, archives or web) and the query form has been classified in the survey. The alternate view from system perspective involves mode of query processing & output visualization for variety of data scopes has been proposed along with it. The output visualization has been characterized as i) Relevance order e.g. Google and Yahoo image search engines ii)time order e.g. Google's Picasa – providing an option for organizing personal collection of images in time order iii) clustered iv) hierarchical and v) composite. Different image similarity comparison measures viz. Euclidian distance, weighted Euclidian, Hausdorff, Mallow, Integrated Region Matching - IRM [Li, 2000] and K –L divergence have been tabulated in the survey.

A. W. Smeulders et al. [Smeulders, 2000] reviewed 200 references in CBIR in year 2000 covering patterns of use, types of pictures, the role of semantics, the sensory gap, color, texture, local geometrical features, accumulative and global features, object and shape features, types and means of feedback and other related issues.

More than 100 articles have been reviewed in a survey on content based multi-media information retrieval by Lew et al. [Lew, 2006] in year 2006. The survey includes review of various face detection techniques for concept detection in simple and complex background. Some of the observations presented in the survey [Lew, 2006] are:

- "One of the most important challenges and perhaps the most difficult problem in semantic understanding of media is visual concept detection in the *presence of complex backgrounds*."
- "Another limiting case is where researchers have examined the problem of detecting visual concepts in laboratory conditions where the background is simple and therefore can be easily segmented."

M¨uller et al. presented analysis on performance evaluation in CBIR and proposed few new measures in [M¨ul, 2001]. They have summarized performance

measures like Precision, Recall, Precision - Recall curves, Precision – Number of Retrievals, Recall – Number of Retrievals, measures specified by TREC (Text REtrieval Conference, co-sponsored by NIST, US) and variational performance measures used by different research-contributors. They have proposed performance evaluation measures similar to those specified by TREC are summarized below:

- $Rank_1$ – Rank at which first relevant image is retrieved
- $\overline{Rank}_1$ - Normalized average rank of relevant images
- PR curves
- P (20), P (50) and P (N_g) – Precision after 20, 50 and N_g relevant images are retrieved, where N_g represents total number of relevant images.
- R_p (0.5) – recall when precision falls below 0.5
- R(100) – Recall after 100 images are retrieved

A comprehensive survey of methods for Colour Image Indexing and Retrieval in Image Databases has been presented by Schettini et al. [Schettini, 2001] covering various color discretization methods, color indexing, issues of histogram comparisons due to color shifts, color spatial indices, illuminant invariant color image indexing techniques and related issues along with methods combining other features with color feature.

2.6 CBIR and Related Techniques

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 and computationally inexpensive 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. The simplest form of histogram is having fixed number of bins. A bin corresponds to a fixed range of intensity values. The range of intensity values defines the width of a bin. Generally, all bins are of equal width. The intensity (color) changes in the same image may shift the bin-membership of pixels altering the intensity distribution.

The nature and amount of change in distribution depends on number of bins and value of change of intensity. The increase in bin-width will reduce the effect of changes in the distribution due to intensity (color) changes. Increase in bin-width results into loss of intra-range distribution information for a larger range. Hence, deciding the bin-width plays important role in the similarity measure methods. Given fixed width of bins may not be suitable to all categories of images. The cumulative histogram gives the cumulative distribution of the intensity values.

The multi-resolution histogram and its use for recognition for image and video retrieval have been proposed in [Hadjidemetriou, 2004]. As defined in it, the multi-resolution histogram is the set of intensity histograms of an image at multiple image resolutions. Like plain histograms, multi-resolution histograms are fast to compute, space efficient, invariant to rigid motions, and robust to noise. In addition, spatial information is directly encoded with multi-resolution histograms. A novel matching algorithm based on the multi-resolution histogram that uses the differences between histograms of consecutive image has been proposed in it [Hadjidemetriou, 2004] along with the effect of shape parameters on the multi-resolution histograms.

Histogram intersection based method for comparing model and image histograms was proposed in [Swain, 1991] for object identification. Histogram refinement based on color coherence vectors was proposed in [Pass, 1996]. 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 [Huang, 1997] which takes 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 [Rao, 1999]. In [Stricker, 1995], 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 [Ma, 2000], 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 [Arbel´aez, 2006]. A generic grouping algorithm based on Oriented Watershed Transform and UCM [Arbel´aez, 2006] has been proposed in [Arbel´aez, 2009] 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. The precise low level processing is very crucial for feature extraction. J. Malik at el. emphasis on perfect boundary detection leading to segmentation and / or object shape description. Local and global cue based contour and junction detection has been proposed in [Maire, 2008].

Image segmentation has been treated as a graph partitioning problem using Normalized cuts in [Shi, 2000]. Two powerful segmentation strategies–mean shift clustering and normalized cuts based accurate and rapid object initialization scheme-weighted mean shift normalized cuts for geodesic active contour model for segmentation of Histopathology images have been presented in [Xu, 2010].

Region based image retrieval, incorporating graphs, multiple low level labels and their propagation, multilevel semantic representation and support vector machine has been proposed by Li et al., implying effectiveness of the method by showing various precision measures only. The recall measure analysis for the incorporated large image

database of Corel consisting of 10000 images would have been helpful throwing lights on effect of various types of feedbacks on region matching. It should be noted that the paper [Li, 2008] uses JSEG [Deng, on line] [Deng, 2001] algorithm for image segmentation.

The brightness and texture gradient based probability of boundary [Martin, 2004] has been used to generate edge-map, which is scale-invariant representation of image from the bottom up, using a piecewise linear approximation of contours and constrained Delaunay triangulation for completing gaps. The curvilinear grouping on top of this graphical/geometric structure using a conditional random field to capture the statistics of continuity and different junction types has been proposed in [Ren, 2008] for contour completion in natural images. A new concept of Boolean derivatives as a fusion of partial derivatives of Boolean function for edge detection algorithms for binary and gray scale image has been presented and results have been compared with traditional edge detection algorithms in [Agaian, 2010]. Color gradient detection based technique for automatic image segmentation has been proposed in [Ugarriza, 2009]. A method for unsupervised determination of hysteresis thresholds for edge detection by combining advantages and disadvantages of thresholding methods by finding best edge map, a subset and an overset of the unknown edge point set has been proposed in [Medina-Carnicer, 2010]. Combined top-down and bottom-up approach for image segmentation has been proposed in [Borenstein, 2008]. Top-down & bottom-up cue based probabilistic method for image segmentation overcoming the limitations of traditional conditional random field (CRF) based approach has been proposed in [Pawan Kumar, 2010]. YCbCr color model based automatic seeded region growing algorithm for image segmentation has been proposed in [Shih, 2005].

YCbCr color space based face detection algorithm for varying lightning conditions and complex background has been proposed in [Hsu, 2002] that incorporates light compensation technique and non-linear color transform. Major face detection techniques have been also listed in [Hsu, 2002]. Analysis and comparison of color representation, color quantization and classification algorithms for skin segmentation have been reported in [Phung, 2005]. Exhaustive survey of face detection issues and techniques has been found in [Yang, 2002]. The foreground objects revealing by separating background in the images of standard data set has been proposed in **[Thakore, 2010, 4]**.

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 [Veltkamp, 2000]. Active contour model – snake has been used in [Kass, 1988] 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. L* u* v* color space recursive mean shift procedure based analysis of multimodal feature space and delineation of arbitrarily shaped cluster can be found in [Comaniciu, 2002]. Scale invariant local shape features with chains of k-connected roughly straight family of contour segments has been used for object class detection in [Ferrari, 2008].

The boundary structures and global shape feature based approach for segmentation and object detection has been proposed in [Toshev, 2010]. Image segmentation and object detection using iterated Graph Cuts, based on local texture features of wavelet coefficient has been reported in [Fukuda, 2008]. The application of watershed algorithm for contour detection leading to segmentation was proposed in [Beucher, 1979].

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 [Rui, 1998], which showed significant improvement in the results. In [Tao, 2008], 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 whereas latter do not. Generalized Bayesian learning framework with target query and a user conception based user-model has been proposed in [Hsu, 2005], 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 [Krishnapuram, 2004] 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 [Malik, 2001] using intervening contour frame work and textons for image segmentation with spectral graph theoretic framework of normalized cuts. As stated in [Malik, 2001] contour based image segmentation approaches have edge detection as the first stage followed by edge linking stage to exploit curvilinear continuity. 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 [Cheng, 2007]. 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 [Chun, 2003] 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 [Saadatmand-Tarzjan, 2007].

A color image edge detection algorithm was proposed in [Dutta, 2009], taking up average maximum color difference value was used to predict the optimum threshold value for a color image and thinning technique was applied to extract proper edges producing comparable results with other edge detection algorithms. The presented method results [Dutta, 2009] were of images containing color patches with no or minimum textures. The performance of the algorithm could have been tested for textured and natural images.

Prasad et al. [Prasad, 2004] proposed image retrieval using integrated color-shape-location index has been proposed based on grouping RGB color space into 25 perceptual color categories, dominant region eccentricity for 8 shape categories and grid cell of image for location identification and indexing with performance measure on database consisting of various national flags and vegetables-fruits images.

The use of CIELab color space based color descriptors for CBIR and comparisons for different quantization methods, histograms calculated using color-only and/or spatial-color information with different similarity measures have been presented in [Gavrielides, 2006] covering retrieval results for images with different transformations like scaling, rotation, cropping, jpeg compression with different quality factors, blurring, illumination changes, contrast adjustments and various adaptive noise attacks.

Morphology-based approaches for CBIR by making use of granulometries independently computed for each sub-quantized color and employing the principle of multi-resolution histograms for describing color, using respectively morphological leveling and watersheds has been proposed in [Aptoula, 2009] for LSH color space.

The relevance feed back based biased discriminative Euclidean embedding (BDEE) was proposed in [Bian, 2010] which parameterizes samples in the original high-dimensional ambient space to discover the intrinsic coordinate of image low-level visual features showing precise modeling of both - the intra-class geometry and interclass discrimination evaluated on Corel image database presenting query response examples.

Comparison of the mean average precision of three content based image retrieval methodologies have been presented in [Vasconcelos, 2007], 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 [Vasconcelos, 2007], 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 [Rasiwasia, 2007], 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 [Rasiwasia, 2008].

Paitakes et al. [Pratikakis, 2006] proposed a novel unsupervised method for image retrieval based on hierarchical watershed algorithm applied on L a* b* color or texture feature space for achieving meaningful segmentation & automatic meaningful region extraction leading to construction of region adjacency graph – RAG incorporating scale based weights in the multi-scale hierarchical frame work for Earth mover's distance (EMD) computation as region-similarity-comparison. The segmentation results produced with their proposed method have been compared with those of JSEG [Deng, on line] [Deng, 2001], E-M algorithm (Blobworld) [Carson, 2002] and graph-based segmentation [Felzenszwalb, 2004]. The image retrieval results have

been compared with the regions generated with their proposed method, JSEG [Deng, on line] [Deng, 2001], E-M algorithm (Blobworld) [Carson, 2002], and graph-based segmentation [Felzenszwalb, 2004] for segmentation by applying their method of region similarity. The mean precision-recall have been measured in [Pratikakis, 2006] for 10 queries per image-class of image database consisting of total 1000 images of 10 different classes with 100 images per class, which reads (approximate values) for all categories of images, highest mean precision of 0.7 at mean recall of 0.07 and highest mean recall of 0.425 with precision of 0.41. Further, P – R curves corresponding to all image categories indicate that it is not possible to retrieve images with precision as 1 at any cost of recall, i.e. for no case, only relevant images (may be very few in number) gets retrieved. As no examples of query responses have been presented in [Pratikakis, 2006], analysis and inference about ordering and ranking of the retrieved resultant images cannot be carried out.

The slope magnitude method along with Sobel, Prewitt, Robert and Canny gradient operators have been used for forming shape image on which block truncation coding (BTC) is applied in [Kekre, 2010, 1] for performing image retrieval on 1000 images of 11 different classes of SIMPLIcity image database [Wang, 2001] [SIMPLIcity, on line]. The performance analysis of 55 queries in a form of precision & recall plotted for number of images retrieved with different combination of methods. Simple morphological edge detection, top-hat morphological edge detection, bottom-hat morphological edge detection methods are combined with BTC for image retrieval in [Kekre, 2010, 2]. The precision is not exceeding 0.8 even for 2 retrieved images. The maximum recall obtained in all seven methods proposed is about 0.35 for 100 retrieved images (relevant + irrelevant) with maximum precision of about 0.35. Here also in both publications [Kekre, 2010, 1] & [Kekre, 2010, 2], no examples of results of query response has been presented to enable analysis of results for order and ranking of retrieved images.

Basak et al. [Basak, 2006] presented Multiple Exemplar-Based Facial Image Retrieval Using Independent Component Analysis as a specific CBIR application for three different image databases, of which one is Caltech [Caltech, on line] [Fei-Fei, 2004] for which, as can be seen from the illustrative query response examples, non-face portion of the image constituting background have been cropped. It is to be noted that our proposed method for similar-face-image retrieval uses Caltech images, where the complex-background, which contributes to major portion of the image has been

excluded based on prominent boundaries and foreground detection based techniques to extract the face region before comparing them for similarity measures.

2.7 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 [QBIC, on line] [Flickner, 1995]. The IBM developed QBIC technology based Ultimedia Manager Product for retrieval of visually similar images [Barber, 1994]. Virage [Virage, on line] 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 [Smith, 1996] [VisualSEEk, on line]. 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 [Photobook, on line]. MARS - Multimedia Analysis and Retrieval Systems [MARS, on line] and FIRE- Flexible Image Retrieval Engine [Fire, on line] 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 [Ma, 1997]. 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 [Ma, 2000] 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 [Laaksonen, 1999]. Visual content descriptors of MPEG-7 (Moving Pictures Expert Group Multimedia Content Description Interface) were used in PicSOM [Laaksonen, 1999] for CBIR techniques and performance comparison with Vector Quantization based system

was proposed in [Laaksonen, 2002]. 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 [Wang, 2001]. The online demo of SIMPLIcity is available at [SIMPLIcity, on line].

A comparative survey of various 42 CBIR systems developed by year 2001 has been reported in [Veltkamp, on line]. The survey compares those systems for mode of querying, features & method of similarity comparisons and indexing techniques. The sample query responses along with URL for demo have been provided (Though very few demo URL pages are accessible as on date, the survey is worth noting for its contents & the query response examples).

Many technology-giants now have research focus on multimedia / video retrieval techniques. The multimedia information retrieval issues and recent research publications covered in the IEEE special issue on Advances in Multimedia Information Retrieval (April 2008) have been summarized in the editorial article [Hanjalic, 2008]. Similarly, the editorial article of special issue on Recent Advances in Image and Video Retrieval, in proceedings of IEE, 2005, summarizes the papers of the issue [O'connor, 2006].

The video-frame based image analysis as the application for video abstraction / video summary has been overviewed in [Li, 2001] along with list of major players in the field.

2.8 Our Observations

The observations learnt and derived for CBIR are as under.

- 'Proper' segmentation is a mandatory requirement for feature extraction.
  - The word 'Proper' is subjective and depends on image characteristics and categories of images. Improper segmentation leads to under-segmentation / over-segmentation of the image under consideration.

Minimization of over segmentation sometimes leads to under segmentation.

- Inter-region and intra-region texture & illumination variations are root causes of over-segmentation.
- The segmentation algorithm should not be very sensitive to such variations.
- The automatic, non-parametric generic segmentation algorithm should be multi-scale / hierarchical to accommodate image characteristic variations.
- The performance of region matching algorithms greatly depends on the quality of segmentation.

o Region matching based techniques are characterized by

- Image comparisons by parts – may not be suitable to some categories of images.
- Finding best matched region by performing one to many region-attribute comparison faces the challenges of resolving clash to determine best matched region and accommodating color / illumination / shape / spatial location changes in the region.

o Proper identification of regions constituting object(s) is the second most important requirement. The process of region-identification may be characterized by region merging / region separating / region propagating / region eliminating operations.

o The user interaction (intervention) in a form of relevance feed-back generally improves the performance by accommodating variations in images and by overcoming limitations of algorithms.

o Stringent image-feature descriptors or strict similarity constraints end up into poor recall with higher precision, assuming other conditions favorable and constant. Broad image-feature descriptors or relaxed similarity measures improve recall by sacrificing precision. (Examples of broad image-descriptors are – histograms, color quantization based features, concept descriptors etc. A Gabor filter applied at various orientations for texture description is an example of stringent image feature.)

o Recall improves if a technique is capable of taking care of variations in color / illumination / poses / shapes in similar images.

- Precision improves if a technique is precise in feature extraction and similarity measures.
- One of the objectives for a design / implementation of a good CBIR system would be 'To retain maximum precision for higher recall in a large image database consisting of variety of images'.
- And, exhaustive testing of methods on variety of image categories is necessary for proving applicability and suitability.

2.9 Our Approaches

The theme of our approaches is *"Relaxed feature description for better Recall and simultaneous emphasizing of reliable processing of cues leading to precise feature extraction for better Precision"*.

Our approaches follow two streams of techniques. The first one is based on broad color feature descriptors called color codes, which is a simple and computationally efficient technique, suitable for image comparison on a broader scale, on the basis of color comparison without considering shapes. The approach is suitable for finding near-similar images having nearly similar color distributions. The second approach emphasizes reliable processing of low level cues for precise and well localized prominent boundaries detection eventually leading to foreground extraction. The extracted foreground is compared on basis of shape - correlation and foreground color codes. The composite approach consisting of foreground shape and foreground color codes provides selectable proportion of weights in composite similarity measures enables users to match the need based on category of query image. The exclusion of background and corresponding features enables object based search for image retrieval. The foreground detection based face extraction method for similar-face-image retrieval from the image containing complex-background has been presented as an application-specific CBIR, illustrating effectiveness of various proposed algorithms.

Our approaches address some of the issues observed by Theo Pavlidis, described in Section 2.3:

- The proposed segmentation method yields good results for wide categories of images enabling foreground (objects) extraction by separating background and facilitating foreground comparison instead of comparison of whole

images. (Addressing of segmentation and object based search issues cited in Section 2.3)

- The foreground extraction method meets the challenges offered by Illumination / pose / viewpoint changes.
- The other approach enables whole image comparison for image retrieval based on broad image-feature descriptors (Color Codes) giving user a selection for foreground or whole image based searches depending upon user's intentions.

Additional characteristics of the approaches are,

- Broader color descriptors – color codes are less sensitive to illumination and color variations up to certain extent, a very helpful characteristic to improve recall.
- The reliable processing yielding precise prominent boundaries & foreground shape combined with broader color descriptors of foreground form a good proposition for object based image retrieval intended for improvements in precision and recall respectively.
- Encouraging results for exhaustive testing of methods on various images.

2.10 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. 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 proposed combination of reliable processing leading to precise feature extraction and broader color descriptors applied to foreground shape leads to encouraging results for foreground based image retrieval for better precision and recall measures. The results of proposed independent approaches, based on – whole image color codes, foreground shape and foreground color codes have shown applicability and suitability of the methods for image retrieval. Various proposed methods put together for development of application specific CBIR - similar-face-image retrieval for images containing complex background, produces results endorsing the effectiveness of methods.

2.11 Concluding Remark

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