Content-Based Image Retrieval - Approaches and Trends of the New Age

Ritendra Datta, Jia Li, and James Z. Wang
The Pennsylvania State University
University Park, PA 16802, USA
datta@cse.psu.edu, jiali@stat.psu.edu, jwang@ist.psu.edu

ABSTRACT

The last decade has witnessed great interest in research on content-based image retrieval. This has paved the way for a large number of new techniques and systems, and a growing interest in associated fields to support such systems. Likewise, digital imagery has expanded its horizon in many directions, resulting in an explosion in the volume of image data required to be organized. In this paper, we discuss some of the key contributions in the current decade related to image retrieval and automated image annotation, spanning 120 references. We also discuss some of the key challenges involved in the adaptation of existing image retrieval techniques to build useful systems that can handle real-world data. We conclude with a study on the trends in volume and impact of publications in the field with respect to venues/journals and sub-topics.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Indexing methods*

General Terms

Algorithms, Documentation, Performance.

Keywords

Content-based image retrieval, annotation.

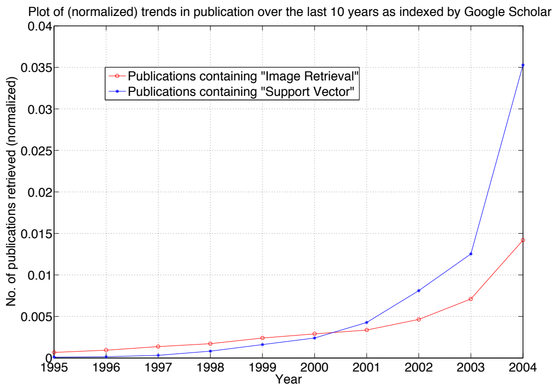
1. INTRODUCTION

Our motivation to organize things is inherent. Over many years we learned that this is a key to progress without the loss of what we already possess. When we started possessing more than what we could manually set to order, we built machines for faster, more efficient or more accurate organization. For decades, *text* in a given language has been set to order, to categorize and to search from, be it manually in the ancient *Bibliotheke*, or automatically as in the modern digital libraries. But when it comes to organizing *images*, man has traditionally outperformed machines for most tasks. One reason which causes this distinction is that text is man's creation, while typical images are a mere replica of what man has seen since birth, the latter being relatively harder to describe concretely. Interpretation of what we see is hard to characterize, and even more so to teach a machine such that any automated organization can be possible. Yet, over the past decade, ambitious attempts have been made to make machines *learn* to understand, index and annotate images representing a wide range of concepts, with much progress.

We, as pursuers of the ultimate goal to build intelligent machines capable of image management the way humans are, feel that it is now time for the community to move aggressively into making it a real-world technology. We sense a paradigm shift in the goals of the next-generation researchers in image retrieval. The need of the hour is to establish how this technology can reach out to the common man in the same way text retrieval techniques have. For example, Google™ and Yahoo!® are household names today, primarily due to the benefits reaped through their use. We envision that image retrieval will enjoy a similar success story if concerted effort is made by the research and user communities in that direction. It is to be noted here the subtle difference in the level of importance of the user community involvement between the text and image retrieval domains, given the same expected level of success. While a text-based search-engine can successfully retrieve documents without understanding the content, there is usually no easy way for a user to give a low-level description of what image she is looking for. Even if she provides an example to search for images with similar content, most current algorithms fail to accurately relate its high-level concept, or the semantics of the image, to its lower level content. The problem with these algorithms is their reliance on visual similarity in judging semantic similarity. Moreover, semantic similarity is a highly subjective measure.

Comprehensive surveys exist on the topic of content-based image retrieval (CBIR) [86, 90], both of which are primarily on publications prior to the year 2000. Surveys also exist on closely related topics such as relevance feedback [119], high-dimensional indexing of multimedia data [9], applications of content-based image retrieval to medicine [74], and applications to art and cultural imaging [15]. One of the reasons for writing this survey is that the field has grown tremendously after 2000, not just in terms of size, but also in the number of new directions explored. To

MIR'05, November 11-12, Singapore, 2005.



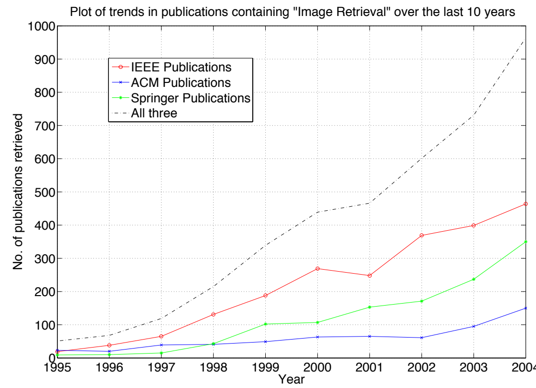


Figure 1: *Top:* Normalized trends in publications having "image retrieval" and "support vector". *Bottom:* Publisher wise break-up of publication count on papers having "image retrieval".

validate this, we conducted a simple test. We searched for publications containing the phrases "Image Retrieval" using Google Scholar [32] and the digital libraries of ACM, IEEE and Springer, within each year from 1995 to 2004. In order to account for (a) the growth of research in computer science as a whole and (b) Google's yearly variations in indexing publications, the Google Scholar results were normalized using the publication count for the word "computer" for that year. A plot on another young and fast-growing field within pattern recognition, support vector machines (SVMs), was generated in a similar manner for comparison. The results can be seen in Fig. 1. Not surprisingly, growth patterns in both these fields are somewhat similar, although SVMs have had a faster growth rate. A precise comment on the growth is not possible using the plots, since there are many implicit assumptions. Nevertheless, the trends in these graphs indicate a roughly exponential growth in interest in image retrieval and closely related topics. We also note that growth in the field has been particularly strong over the last five years, spanning new techniques, new support systems, and diverse application domains. Yet, a brief scanning of about 300 relevant papers published in the last five years revealed that less than 20% were concerned with applications or real-world systems. This may not be a cause for concern, since the theoretical foundation (if any such exists) behind how we humans interpret images is still an open problem. But then, with hundreds of different approaches proposed so far, and no consensus reached on any, it is rather optimistic to believe that we will chance upon a reliable one in the near future. Instead, it may make more sense to build systems that are useful, even if their use is limited to specific domains. A way to see this is that natural language interpretation is an unsolved problem, yet text-based search engines have proved very useful.

In this paper, we review recent work (i.e., year 2000 onwards) in automatic image retrieval and annotation, with a focus on real-world usage and applications of the same. We leave out retrieval from video sequences and text caption based image search from our discussion. The rest of this paper is arranged as follows: Some of the key ideas behind the proposed approaches are discussed in Sec. 2. This leads us to a discussion on some of the most desirable features of real-world image indexing systems, in Sec. 3, learned through our own experiences with retrieval system implementations. A discussion on the publication trends within the field with respect to venues/journals and subtopics is presented in Sec. 4. We conclude in Sec. 5.

2. NEW IDEAS AND APPROACHES

We do not yet have a universally acceptable algorithmic means of characterizing human vision, more specifically in the context of image understanding. Hence it is not surprising to see continuing efforts towards it, either building up on prior work [90] or exploring novel directions. In this section, we discuss recent literature on some key aspects of content-based image retrieval and automated annotation.

2.1 Feature Extraction

Most systems perform feature extraction as a pre-processing step, obtaining global image features like color histogram or local descriptors like shape and texture. A region based dominant color descriptor indexed in 3-D space along with their percentage coverage within the regions is proposed in [25], and shown to be more computationally efficient in similarity based retrieval than traditional color histograms. The authors argue that this compact representation is more efficient than high-dimensional histograms in terms of search and retrieval, and it also gets around some of the drawbacks associated with earlier propositions such as dimension reduction and color moment descriptors. In [35], a multi-resolution histogram capturing spatial image information has been shown to be effective in retrieving textured images, while retaining the typical advantages of histograms. In [46], Gaussian mixture vector quantization (GMVQ) is used to extract color histograms and is shown to yield better retrieval than uniform quantization and vector quantization with squared error. A set of color and texture descriptors rigorously tested for inclusion in the MPEG-7 standard, and well suited to natural images and video, is described in [69]. These include histogram-based descriptors, dominant color descriptors, spatial color descriptors and texture descriptors suited for browsing and retrieval. Texture features have been modeled on the marginal distribution of wavelet coefficients using generalized Gaussian distributions, in [26].

Shape is a key attribute of segmented image regions, and its efficient and robust representation plays an important role in retrieval. A shape similarity measure using discrete curve evolution to simplify contours, is discussed in [59]. Doing this contour simplification helps to remove noisy and irrelevant shape features from consideration. A new shape descriptor for shape matching, referred to as *shape context*, has been proposed [7] which is fairly compact yet robust to a number of geometric transformations. A dynamic programming (DP) approach to shape matching has been proposed in [81]. One problem with this approach is that computation of Fourier descriptors and moments is slow, although pre-computation may help produce real-time results. Continuing with Fourier descriptors, exploitation of both the amplitude and phase and using Dynamic Time Warping (DTW) distance instead of Euclidean distance has been shown to be an accurate shape matching technique in [6]. The rotational and starting point invariance otherwise obtained by discarding the phase information is maintained here by adding compensation terms to the original phase, thus allowing its exploitation for better discrimination.

For characterizing shape within images, reliable segmentation is critical, without which the shape estimates are largely meaningless. Even though the general problem of segmentation in the context of human perception is far from being solved, there have been some interesting new directions, one of the most important being segmentation based on the Normalized Cuts criteria [89]. This approach, based primarily on the theory of spectral clustering, has been extended to textured image segmentation by using cues of contour and texture differences [68], and to incorporate partial grouping priors into the segmentation process by solving a constrained optimization problem [113]. The latter has potential for incorporating real-world application-specific priors, e.g. location and size cues of organs in pathological images. Talking of medical imaging, 3D brain magnetic resonance (MR) images have been segmented using Hidden Markov Random Fields and the Expectation-Maximization (EM) algorithm [115], and the spectral clustering approach has found some success in segmenting vertebral bodies from sagittal MR images [10]. Among other recent approaches proposed are segmentation based on the mean shift procedure [20], multi-resolution segmentation of low depth of field images [103], a Bayesian framework based segmentation involving the Markov chain Monte Carlo technique [97], and an EM algorithm based segmentation using a Gaussian mixture model [12], forming *blobs* suitable for image querying and retrieval. A sequential segmentation approach that starts with texture features and refines segmentation using color features is explored in [16].

While there is no denying that achieving good segmentation is a big step forward in image understanding, some of the issues plaguing current techniques are speed considerations, reliability of good segmentation, and a robust and acceptable benchmark for assessment of the same. In the case of image retrieval, some of the ways of getting around this problem have been to reduce dependence on reliable segmentation [12], to involve every generated segment of an image in the matching process to obtain *soft* similarity measures [104], or to characterize spatial arrangement of color and texture using block-based multi-resolution hidden Markov models [63, 65], a technique that has been extended to segment 3D volume images as well [64]. Another alternative has been to use principles of perceptual grouping to hierarchically extract image structure [44].

Features based on local invariants such as *corner points* or *interest points* that have traditionally been used for stereo matching are being used extensively in image retrieval. Scale and affine invariant interest points that can deal with significant affine transformations and illumination changes have been shown as effective features for image retrieval [70]. In similar lines, wavelet-based *salient points* have been used for retrieval [93]. The significance of such special points lie in their compact representation of important image regions, leading to efficient indexing and good discriminative power, especially in object-based retrieval. A discussion on the pros and cons of different types of color interest points used in image retrieval can be found in [33], while a comparative performance evaluation of the various proposed interest point detectors is reported in [71].

The selection of appropriate features for content-based image retrieval and annotation systems remain largely ad-hoc, with some exceptions. One heuristic in the selection process is to have application-specific feature sets. Although semantics-sensitive feature selection has been shown effective in image retrieval [104], the need for a uniform feature space for efficient search and indexing limits heterogeneous feature set size to some extent. When a large number of image features are available, one way to improve generalization and efficiency in classification and indexing is to work with a feature subset. To avoid a combinatorial search, an automatic feature subset selection algorithm for SVMs has been proposed in [107]. Some of the other recent, more generic feature selection propositions involve boosting [94], evolutionary searching [54], Bayes classification error [11], and feature dependency/similarity measures [72]. A survey and performance comparison of some recent algorithms on the topic can be found in [34].

2.2 Approaches to Retrieval

Once a decision on the visual feature set choice has been made, how to steer them towards accurate image retrieval is the next concern. There has been a large number of fundamentally different frameworks proposed in the last few years. Leaving out those discussed in [90], here we briefly talk about some of the more recent approaches.

A semantics-sensitive approach to content-based image retrieval has been proposed in [104]. A semantic categorization (e.g., graph - photograph, textured - non-textured) for appropriate feature extraction followed by a region based overall similarity measure, allows robust image matching. An important aspect of this system is its retrieval speed. The matching measure, termed integrated region matching (IRM), has been constructed for faster retrieval using region feature clustering and the most similar highest priority (MSHP) principle [28]. Region based image retrieval has also been extended to incorporate spatial similarity using the Hausdorff distance on finite sized point sets [55], and to employ fuzziness to characterize segmented regions for the purpose of feature matching [17]. A framework for region-based image retrieval using region codebooks and learned region weights has been proposed in [49]. A new representation for object retrieval in cluttered images without relying on accurate segmentation has been proposed in [3]. Another perspective in image retrieval has been region-based querying using homogeneous color-texture segments called *blobs*, instead of image to image matching [12]. For example, if one or more segmented blobs are identified by the user as roughly corresponding to the concept "tiger", then her search can comprise of looking for a tiger within other images, possibly with varying backgrounds. While this can lead to a semantically more precise representation of the user's query objects in general, it also requires greater involvement from and dependence on her. For finding images containing scaled or translated versions of query objects, retrieval can also be performed without the user's explicit region labeling [76].

Instead of using image segmentation, one approach to retrieval has been the use of hierarchical perceptual grouping of primitive image features and their inter-relationships to characterize structure [44]. Another proposition has been the use of vector quantization (VQ) on image blocks to generate *codebooks* for representation and retrieval, taking inspiration from data compression and text-based strategies [120]. A windowed search over location and scale has been shown more effective in object-based image retrieval than methods based on inaccurate segmentation [42]. A hybrid approach involves the use of rectangular blocks for coarse foreground/background segmentation on the user's query region-of-interest (ROI), followed by the database search using only the foreground regions [24]. For textured images, segmentation is not critical. A method for texture retrieval by a joint modeling of feature extraction and similarity measurement using the Kullback-Leibler distance for statistical model comparison has been proposed in [26]. Another wavelet-based retrieval method involving salient points has been proposed in [93]. Fractal block code based image histograms have been shown effective in retrieval on textured image databases [82]. The use of the MPEG-7 content descriptors to train self-organizing maps (SOM) for the purpose of image retrieval has been explored in [57].

Among other new approaches, anchoring-based image retrieval system has been proposed in [77]. Anchoring is based on the fairly intuitive idea of finding a set of representative "anchor" images and deciding semantic proximity between an arbitrary image pair in terms of their similarity to these anchors. Despite the reduced computational complexity, the relative image distance function is not guaranteed to be a metric. For similar reasons, a number of approaches have relied on the assumption that the image feature space is a manifold embedded in Euclidean space [38, 101, 39]. Clustering has been applied to image retrieval to help improve interface design, visualization, and result pre-processing [19, 61, 116]. A statistical approach involving the Wald-Wolfowitz test for comparing non-parametric multivariate distributions has been used for color image retrieval [92], representing images as sets of vectors in the RGB-space. Multiple-instance learning was introduced to the CBIR community in [114].

A number of probabilistic frameworks for image retrieval have been proposed in the last few years [48, 102]. The idea in [102] is to integrate feature selection, feature representation, and similarity measure into a combined Bayesian formulation, with the objective of minimizing the probability of retrieval error. One problem with this approach is the computational complexity involved in estimating probabilistic similarity measures. Using VQ to approximately model the probability distribution of the image features, the complexity is reduced [99], making the measures more practical for real-world systems.

2.3 Annotation and Concept Detection

While image retrieval has been active over the years, an emerging new and possibly more challenging field is automatic concept recognition from visual features of images. The challenge is primarily due to the *semantic gap* [90] that exists between low level visual features and high level concepts. A note on the topic of concept and annotation: The primary purpose of a practical content-based image retrieval system is to discover images pertaining to a given concept in the absence of reliable meta-data. All attempts at automated concept discovery, annotation, or linguistic indexing essentially adhere to that objective more closely than do systems which return an ordered set of similar images. Of course, ranked results have their own role to play, e.g. visualization of search results, retrieval of specific instances within a semantic class of images etc.

Annotation, on the other hand, allows for image search through the use of text. For this purpose, automated annotation tends to be more practical for large data sets than a manual process. If the resultant automated mapping between images and words can be trusted, then text-based image searching can be semantically more meaningful than CBIR. Image understanding has been attempted through automated concept detection. The annotation process can be thought of as a subset of concept detection, i.e., images pertaining to the same concept can be described linguistically in different ways based on the specific instance of the concept. The question is whether visual features of images convey anything about their concept or not.

Concept detection through supervised classification, involving simple concepts such as city, landscape, sunset, and forest, have been achieved with high accuracy in [98]. An extension of multiple-instance learning has been shown effective for categorization of images into semantic classes [18]. Learning concepts from user's feedback and within a dynamically changing image database using Gaussian mixture models is discussed in [27]. An approach to *soft* annotation, using Bayes Point machines, to give images a confidence level for each trained semantic label has been explored in [14]. This vector of confidence labels can then be exploited to rank relevant images in case of a keyword search. Automated annotation of pictures with a few hundreds of words using two-dimensional multi-resolution hidden Markov models has been explored in [65]. While the classification process chooses a set of categories an image may belong to, the annotation set is chosen in a way that favors statistically salient words for a given image. A confidence based dynamic ensemble of SVM classifiers has been used for the purpose of annotation in [62].

Many of the approaches to image annotation have been inspired by research in the text domain. In [29], the problem of annotation is treated as a *translation* from a set of image segments to a set of words, in a way analogous to linguistic translation. Hierarchical statistical methods for modeling the association between image segments and words, for the purpose of automated annotation, have been proposed in [5, 8]. Generative language models have been used for the task of image annotation in [45, 60]. Closely related is an approach, involving coherent language models, which exploits word-to-word correlations to strengthen annotation decisions [47]. All the annotation strategies discussed so far model visual and textual features separately prior to association. A departure from this trend is seen in [73], where latent semantic analysis (LSA) is used on uniform vectored data consisting of both visual features and textual annotations. The LSA model, previously used in document analysis, helps to identify semantically meaningful subspaces in the visual-textual feature space.

Automated image annotation is a difficult question. We humans segment objects better than machines, having learned to associate over a long period of time, through multiple viewpoints, and literally through a "streaming video" at all times, which partly accounts for our natural segmentation. The association of words and *blobs* become truly meaningful only when blobs isolate objects well. Moreover, how exactly our brain does this association is still unclear. While Biology tries to answer this fundamental question, researchers in information retrieval tend to take a pragmatic stand in that they aim to build retrieval and annotation systems that have practical significance.

2.4 Relevance Feedback and Learning

Relevance feedback (RF) is a query modification technique, originating in information retrieval, that attempts to capture the user's precise needs through iterative feedback and query refinement. Ever since its inception in the image retrieval community [87], a great deal of interest has been generated. In the absence of a reliable framework for characterizing high-level semantics of images and human subjectivity of perception, the user's feedback provides a way to learn case-specific query semantics. We present short overview of recent progress in RF. A more complete review can be found in [119].

Normally, user's RF results in only a small number of labeled images pertaining to each high level concept. Learning based approaches are typically used to appropriately modify the feature set or the similarity measure. To circumvent the problem of learning from small training sets, a discriminant-EM algorithm has been proposed to make use of unlabeled images in the database for selecting more discriminating features [110]. Learning from RF in the case of systems that compare images using multiple visual features has been studied and an optimized learning strategy suggested in [85]. Methods for performing RF using the visual features as well as associated keywords (semantics) in unified frameworks have been reported in [67, 117]. One problem with RF is that after every round of user interaction, usually the top results with respect to the query have to be recomputed using a modified similarity measure. A way to speed up this *nearest-neighbor* search has been proposed in [109]. Another issue is the user's patience in supporting multi-round feedbacks. A way to reduce the user's interaction is to incorporate logged feedback history into the current query [41]. History of usage can also help in capturing the relationship between high level semantics and low level features [36]. We can also view RF as an *active learning* process, where the learner chooses an appropriate subset for feedback from the user in each round based on her previous rounds of feedback, instead of choosing a random subset. Active learning using SVMs was introduced into the field of image retrieval in [95]. Extensions to the active learning process have also been proposed [31, 37].

With the increase in popularity of region-based image retrieval [12, 104], attempts have been made to incorporate the *region* factor into RF using query point movement and support vector machines [49, 50]. A tree-structured self-organizing map has been used as an underlying technique for RF [58] in a content-based image retrieval system [57]. Probabilistic approaches have been taken in [21, 91, 100]. A clustering based approach to RF incorporating the user's perception in case of complex queries has been studied in [53]. In [39], manifold learning on the user's feedback based on geometric intuitions about the underlying feature space is proposed. While most RF algorithms proposed deal with a two-class problem, i.e., relevant or irrelevant images, another way of looking at RF is to consider multiple relevant and irrelevant groups of images using an appropriate user interface [40, 79, 118]. For example, if the user is looking for *cars*, then she can highlight groups of *blue cars* and *red cars* as relevant examples, since it may not be possible to represent the concept *car* uniformly in any low level feature space. Yet another deviation from norm is the use of multi-level relevance scores to incorporate the relative degrees of relevance of certain images to the user's query [108].

2.5 Hardware and Interface Support

Real-world applications often demand real-time response. One way to make the increasingly complex image retrieval algorithms practical is to use domain-specific hardware acceleration. Unfortunately, very little has been explored in this direction. The notable few include an FPGA implementation of a color histogram based image retrieval system [56], an FPGA implementation for sub-image retrieval within an image database [78], and a method for efficient retrieval in a network of imaging devices [106].

While the focus has generally been on retrieval and annotation performance, presentation of results has often taken a back-seat. Subjectivity in the needs as well as interpretation of results is an issue. One way around it is to allow for greater flexibility in querying/visualization. Some recent innovations in querying include sketch-based retrieval of color images [13], querying using 3-D models [4] motivated by the fact that 2-D image queries are unable to capture the spatial arrangement of objects within the image, and a multi-modal system involving hand-gestures and speech for querying and RF [52]. For image annotation systems, one way to conveniently create sufficiently representative manually annotated training databases is by building interactive, public domain games [1].

For designing querying/visualization for image retrieval systems, it helps to understand factors like how people manage their digital photographs [84] or frame their queries for visual art images [23]. In [83], user studies on various ways of arranging images for browsing purposes are conducted, and the observation is that both visual feature based arrangement and concept-based arrangement have their own merits and demerits. Thinking beyond the typical grid-based arrangement of top matching images, spiral and concentric visualization of retrieval results have been explored in [96]. Efficient ways of browsing large images interactively, e.g., those encountered in pathology or remote sensing, using small displays over a communication channel are discussed in [66]. Speaking of small displays, user log based approaches to smarter ways of image browsing on mobile devices have been proposed in [111]. For personal images, innovative arrangements of query results based on visual content, time-stamps, and efficient use of screen space add new dimensions to the browsing experience [43].

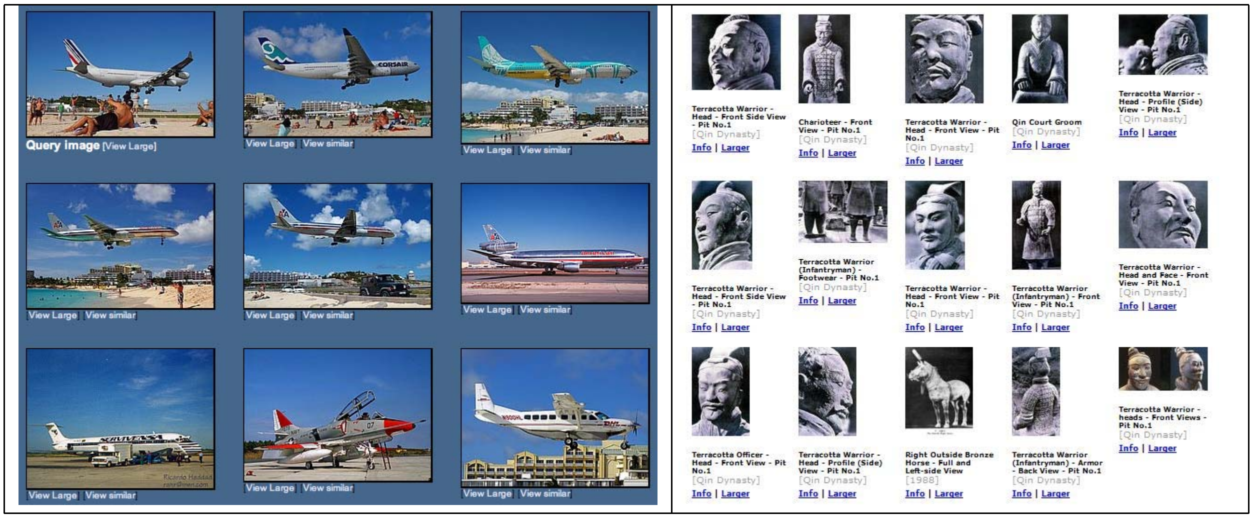


Figure 2: *Left:* Image search on Airliners.net. *Right:* Searching art images in Global Memory Net.

3. REAL-WORLD REQUIREMENTS

Building real-world systems involve regular user feedback during the development process, as required in any other software development life cycle. Not many image retrieval systems are deployed for public usage, save for Google Images or Yahoo! Images (which are based primarily on surrounding meta-data rather than content). There are, however, a number of propositions for real-world implementation. For brevity of space we are unable to discuss them in details, but it is interesting to note that CBIR has been applied to fields as diverse as Botany, Astronomy, Mineralogy, and Remote sensing [105, 22, 80, 88]. With so much interest in the field at the moment, there is a good chance that CBIR based real-world systems will diversify and expand further. We have implemented an IRM-based [104] publicly available similarity search tool on an on-line database of over $800,000$ airline-related images [2]. Another on-going project is the integration of similarity search functionality to a large collection of art and cultural images [30]. Screen-shots can be seen in Fig. 2. Based on our experience with implementing CBIR systems on real-world data for public usage, we list here some of the issues that we found to be critical for real-world deployment.

Performance: The most critical issue is the quality of retrieval and how relevant it is to the domain-specific user community. Most of the current effort is concentrated on improving performance in terms of their precision and recall.

Semantic learning: To tackle the problem of semantic gap faced by CBIR, learning image semantics from training data and developing retrieval mechanisms to efficiently leverage semantic estimation are important directions.

Volume of Data: Public image databases tend to grow into unwieldy proportions. The software system must be able to efficiently handle indexing and retrieval at such scale.

Heterogeneity: If the images originate from diverse sources, parameters such as quality, resolution and color depth are likely to vary. This in turn causes variations in color and texture features extracted. The systems can be made more robust by suitably tackling these variations.

Concurrent Usage: In on-line image retrieval systems, it is likely to have multiple concurrent users. While most systems have high resource requirements for feature extraction, indexing etc., they must be efficiently designed so as not to exhaust the host server resources. Alternatively, a large amount of resources must be allocated.

Multi-modal features: The presence of reliable meta-data such as audio or text captions associated with the images can help understand the image content better, and hence leverage the retrieval performance. On the other hand, ambiguous captions such as "wood" may actually add to the confusion, in which case the multi-modal features together may be able to resolve the ambiguity.

User-interface: As discussed before, a greater effort is needed to design intuitive interfaces for image retrieval such that people are actually able to use the tool to their benefit.

Operating Speed: Time is critical in on-line systems as the response time needs to be low for good interactivity. Implementation should ideally be done using efficient algorithms, especially for large databases. For computationally complex tasks, off-line processing and caching the results in parts is one possible way out.

System Evaluation: Like any other software system, image retrieval systems are also required to be evaluated to test the feasibility of investing in a new version or a different product. The design of a CBIR benchmark requires careful design in order to capture the inherent subjectivity in image retrieval. One such proposal can be found in [75].

4. CURRENT RESEARCH TRENDS

We briefly analyzed publication trends in image retrieval and annotation since the year 2000. We used Google Scholar for this purpose. We queried on the phrase "image OR images OR picture OR pictures OR content-based OR indexing OR 'relevance feedback' OR annotation ", year 2000 onwards, for publications in the journals - IEEE T. Pattern Analysis and Machine Intelligence (PAMI), IEEE T. Image Processing (TIP), IEEE T. Circuits and Systems for Video Technology (CSVT), IEEE T. Multimedia (TOM), J. Machine Learning Research (JMLR), International J. Computer Vision (IJCV), Pattern Recognition Letters (PRL), and ACM Computing Surveys (SURV) and conferences - IEEE Computer Vision and

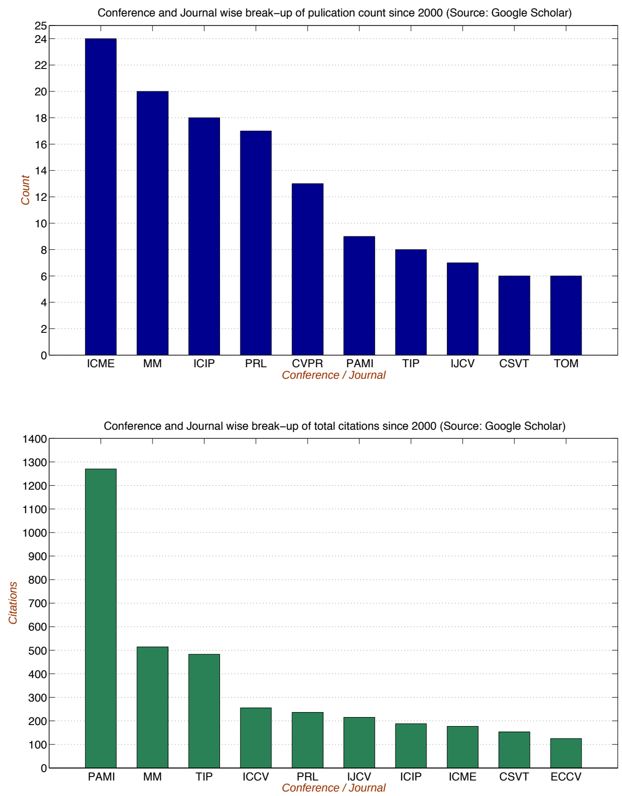


Figure 3: Conference/Journal wise publication statistics on topics closely related to image retrieval, 2000 onwards. *Top:* Publication counts. *Bottom:* Total citations.

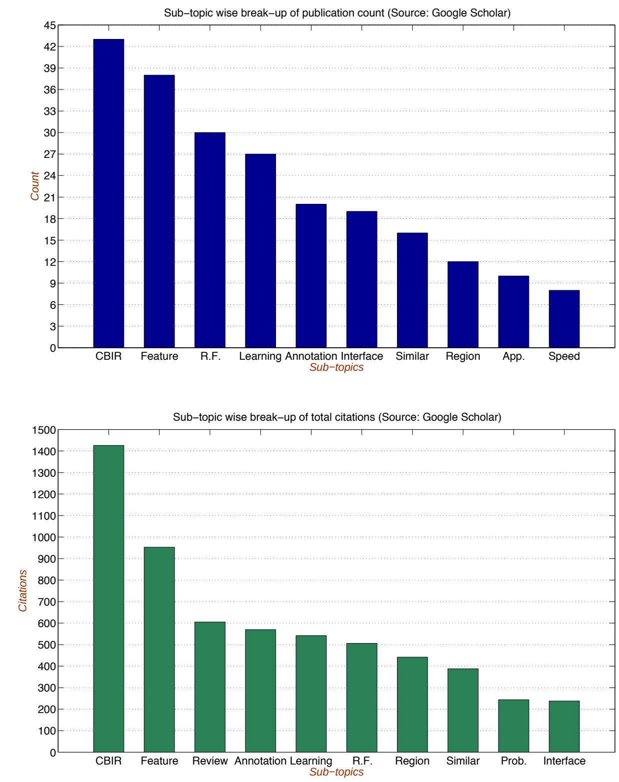


Figure 4: Publication statistics on sub-topics of image retrieval, 2000 onwards. *Top:* Publication Counts. *Bottom:* Total citations. *Abbreviations: Feature* - Feature Extraction, *R.F.* - Relevance Feedback, *Similar* - Image similarity measures, *Region* - Region based approaches, *App.* - Applications, *Prob.* - Probabilistic approaches, *Speed* - Speed and other performance enhancements.

Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV), European Conference on Computer Vision (ECCV), IEEE International Conference on Image Processing (ICIP), ACM Multimedia (MM), ACM SIG Information Retrieval (IR), and ACM Human Factors in Computing Systems (CHI). Relevant papers among the top 100 results in each of these searches were used for the study. Google Scholar presents results roughly in decreasing order of citations (again, only rough approximations to the actual numbers). Limiting search to the top few papers translates to reporting statistics on work with noticeable impact. We gathered statistics on two parameters, (1) publishing venue/journal, and (2) sub-topics of interest. These trends are reported in terms of (a) number of papers, and (b) total number of citations. Plots of these scores are presented in Fig. 3 and Fig. 4. Note that the tabulation is not mutually exclusive (i.e. one paper can have contributions in multiple different sub-topics such as 'Learning' and 'Relevance Feedback', and hence are counted under both heads), neither is it exhaustive or scientifically precise (Google's citation values may not be accurate). Nevertheless, these plots convey general trends in the relative impact of scholarly work. Readers are advised to use their discretion when interpreting these results.

5. CONCLUSIONS

We have presented a brief survey on work related to the young and exciting fields of content-based image retrieval and automated image annotation, spanning 120 publications in the current decade. We believe that the field will experience a paradigm shift in the foreseeable future, with the focus being more on application-oriented, domain-specific work, generating considerable impact in day-to-day life. We have laid out some guidelines for building practical, real-world systems that we perceived during our own implementation experiences. Finally, we have compiled research trends in CBIR and automated annotation using Google Scholar's search tool and citation scores. The trends indicate that while systems, feature extraction, and relevance feedback have received a lot of attention, application-oriented aspects such as interface, visualization, scalability, and evaluation have traditionally received lesser consideration. We feel that for all practical purposes, these aspects should also be considered equally important. Meanwhile, the quest for robust and reliable image understanding technology needs to continue as well. The future of this field depends on the collective focus and overall progress in each aspect of image retrieval, and how much the ordinary individual stands to benefit from it.

6. ACKNOWLEDGMENTS

The material is based upon work supported by the US National Science Foundation under Grant Nos. IIS-0219272 and IIS-0347148, the PNC Foundation, and SUN Microsystems. Conversations with Dhiraj Joshi have been helpful. `Riemann.ist.psu.edu` provides more information.

7. REFERENCES

[1] L. von Ahn and L. Dabbish, "Labeling Images with a Computer Game," *Proc. ACM CHI*, 2004.

[2] Airliners.net, `http://www.airliners.net`.

[3] J. Amores, N. Sebe, P. Radeva, T. Gevers, and A. Smeulders, "Boosting Contextual Information in Content-Based Image Retrieval," *Proc. Workshop on Multimedia Information Retrieval, ACM Multimedia*, 2004.

[4] J. Assfalg, A. Del Bimbo, and P. Pala, "Three-Dimensional Interfaces for Querying by Example in Content-Based Image Retrieval," *IEEE Trans. Visualization and Computer Graphics*, 8(4):305–318, 2002.

[5] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching Words and Pictures," *Journal of Machine Learning Research*, 3:1107-1135, 2003.

[6] I. Bartolini, P. Ciaccia, and M. Patella, "WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(1):142–147, 2005.

[7] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[8] D. M. Blei and M. I. Jordan, "Modeling Annotated Data," *Proc. ACM Conference on Research and Development in Information Retrieval*, 2003

[9] C. Bohm, S. Berchtold, and D. A. Keim, "Searching in High-Dimensional SpacesIndex Structures for Improving the Performance of Multimedia Databases," *ACM Computing Surveys*, 33(3):322-373, 2001.

[10] J. Carballido-Gamio, S. Belongie, and S. Majumdar "Normalized Cuts in 3-D for Spinal MRI Segmentation," *IEEE Trans. Medical Imaging*, 23(1):36–44, 2004.

[11] G. Carneiro and N. Vasconcelos, "Minimum Bayes Error Features for Visual Recognition by Sequential Feature Selection and Extraction," *Proc. Canadian Conference on Computer and Robot Vision*, 2005.

[12] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-maximization and Its Application to Image Querying," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(8):1026-1038, 2002.

[13] A. Chalechale, G. Naghdy, and A. Mertins, "Sketch-Based Image Matching Using Angular Partitioning," *IEEE Trans. Systems, Man, and Cybernetics*, 35(1):28–41, 2005.

[14] E. Y. Chang, K. Goh, G. Sychay, and G. Wu, "CBSA: Content-based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines," *IEEE Trans. Circuits and Systems for Video Technology*, 13(1):26–38, 2003.

[15] C.-c. Chen, H. Wactlar, J. Z. Wang, and K. Kiernan, "Digital Imagery for Significant Cultural and Historical Materials - An Emerging Research Field Bridging People, Culture, and Technologies," *International Journal on Digital Libraries*, 5(4):275–286, 2005.

[16] J. Chen, T.N. Pappas, A. Mojsilovic, and B. Rogowitz, "Adaptive image segmentation based on color and texture," *Proc. IEEE International Conference on Image Processing*, 2002.

[17] Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):252–1267, 2002.

[18] Y. Chen and J. Z. Wang, "Image Categorization by Learning and Reasoning with Regions," *Journal of Machine Learning Research*, 5:913-939, 2004.

[19] Y. Chen, J. Z. Wang, and R. Krovetz, "CLUE: Cluster-Based Retrieval of Images by Unsupervised Learning," *IEEE Trans. Image Processing*, 14(8):1187–1201, 2005.

[20] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5), 603–619, 2002.

[21] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments," *IEEE Trans. Image Processing*, 9(1):20–37, 2000.

[22] A. Csillaghy, H. Hinterberger, and A.O. Benz, "Content Based Image Retrieval in Astronomy," *Information Retrieval*, 3(3):229–241, 2000.

[23] S. J. Cunningham, D. Bainbridge, and M. Masoodian, "How People Describe Their Image Information Needs: A Grounded Theory Analysis of Visual Arts Queries," *Proc. ACM/IEEE Joint Conference on Digital Libraries*, 2004.

[24] C. Dagli and T. S. Huang, "A Framework for Grid-Based Image Retrieval," *Proc. IEEE International Conference on Pattern recognition*, 2004.

[25] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and H. Shin, "An Efficient Color Representation for Image Retrieval," *IEEE Trans. Image Processing*, 10(1):140–147, 2001.

[26] M. N. Do and M. Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance," *IEEE Trans. Image Processing*, 11(2):146–158, 2002.

[27] A. Dong and B. Bhanu, "Active Concept Learning for Image Retrieval in Dynamic Databases," *Proc. IEEE International Conference on Computer Vision*, 2003.

[28] Y. Du and J. Z. Wang "A Scalable Integrated Region-Based Image Retrieval System," *Proc. IEEE International Conference on Image Processing*, 2001.

[29] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In Seventh European Conference on Computer Vision, pages 97–112, 2002.

[30] Global Memory Net, `http://www.memorynet.org`.

[31] K.-S. Goh, E. Y. Chang, and W.-C. Lai, "Multimodal Concept-Dependent Active Learning for Image Retrieval," *ACM Multimedia*, 2004.

[32] Google Scholar, `http://scholar.google.com`.

[33] V. Gouet and N. Boujemaa, "On the Robustness of Color Points of Interest for Image Retrieval" *Proc. IEEE International Conference on Image Processing*, 2002.

[34] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, 3:1157–1182, 2003.

[35] E. Hadjidemetriou, M. D. Grossberg, and S. K. Nayar, "Multiresolution Histograms and Their Use for Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(7):831–847, 2004.

[36] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, "A Memory Learning Framework for Effective Image Retrieval," *IEEE Trans. Image Processing*, 14(4):511–524, 2005.

[37] J. He, M. Li, H.-J. Zhang, H. Tong, C. Zhang, "Mean Version Space: a New Active Learning Method for Content-Based Image Retrieval," *Proc. Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia*, 2004.

[38] X. He, "Incremental Semi-Supervised Subspace Learning for Image Retrieval," *Proc. ACM Multimedia*, 2004.

[39] X. He, W.-Y. Ma, and H.-J. Zhang, "Learning an Image Manifold for Retrieval," *Proc. ACM Multimedia*, 2004.

[40] C.-H. Hoi and M. R. Lyu, "Group-based Relevance Feedback with Support Vector Machine Ensembles," *Proc. International Conference on Pattern Recognition*, 2004.

[41] C.H. Hoi and M. R. Lyu, "A Novel Logbased Relevance Feedback Technique in Contentbased Image Retrieval," *Proc. ACM Multimedia*, 2004.

[42] D. Hoiem, R. Sukthankar, H. Schneiderman, and L. Huston, "Object-Based Image Retrieval Using the Statistical Structure of Images," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[43] D. F. Huynh, S. M. Drucker, P. Baudisch, and C. Wong, "Time Quilt: Scaling up Zoomable Photo Browsers for Large, Unstructured Photo Collections," *Proc. ACM CHI*, 2005.

[44] Q. Iqbal and J. K. Aggarwal, "Retrieval by Classification of Images Containing Large Manmade Objects Using Perceptual Grouping," *Pattern Recognition Journal*, 35(7):1463–1479, 2002.

[45] J. Jeon, V. Lavrenko, and R. Manmatha, "Automatic Image Annotation and Retrieval using Cross-media Relevance Models", *Proc. ACM Conference on Research and Development in Information Retrieval*, 2003.

[46] S. Jeong, C. S. Won, and R.M. Gray, "Image retrieval using color histograms generated by Gauss mixture vector quantization," *Computer Vision and Image Understanding*, 9(1–3):44–66, 2004.

[47] R. Jin, J. Y. Chai, and L. Si, "Effective Automatic Image Annotation Via A Coherent Language Model and Active Learning," *Proc. ACM Multimedia*, 2004.

[48] R. Jin and A.G. Hauptmann, "Using a Probabilistic Source Model for Comparing Images," *Proc. IEEE International Conference on Image Processing*, 2002.

[49] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "An Efficient and Effective Region-Based Image Retrieval Framework," *IEEE Trans. Image Processing*, 13(5):699–709, 2004.

[50] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "Relevance Feedback in Region-Based Image Retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, 14(5):672–681, 2004.

[51] D. Joshi, J. Z. Wang, and J. Li, "The Story Picturing Engine: Finding Elite Images to Illustrate a Story Using Mutual Reinforcement," *Proc. Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia*, 2004.

[52] T. Kaster, M. Pfeiffer, and C. Bauckhage, "Combining Speech and Haptics for Intuitive and Efficient Navigation through Image Databases," *Proc. International Conference on Multimodal Interfaces*, 2003.

[53] D.-H. Kim and C.-W. Chung,"Qcluster: Relevance Feedback Using Adaptive Clustering for Content Based Image Retrieval," *Proc. ACM Conference on Management of Data*, 2003.

[54] Y. S. Kim, W. N. Street, and F. Menczer, "Feature Selection in Unsupervised Learning via Evolutionary Search," *Proc. ACM International Conference on Knowledge Discovery and Data Mining*, 2000.

[55] B. Ko and H. Byun, "Integrated Region-Based Image Retrieval Using Region's Spatial Relationships," *Proc. IEEE International Conference on Pattern Recognition*, 2002.

[56] L. Kotoulas and I. Andreadis, "Colour Histogram Content-based Image Retrieval and Hardware Implementation," *IEE Proc. Circuits, Devices and Systems*, 150(5):387–393, 2003.

[57] J. Laaksonen, M. Koskela, and E. Oja, "PicSOM-Self-Organizing Image Retrieval With MPEG-7 Content Descriptors," *IEEE Trans. Neural Networks*, 13(4):841–853, 2002.

[58] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "Self-Organising Maps as a Relevance Feedback Technique in Content-Based Image Retrieval," *Pattern Analysis and Applications*, 4:140-152, 2001.

[59] L.J. Latecki and R. Lakamper, "Shape Similarity Measure Based on Correspondence of Visual Parts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1185–1190, 2000.

[60] V. Lavrenko, R. Manmatha, and J. Jeon, "A Model for Learning the Semantics of Pictures," *Proc. Advances in Neutral Information Processing Systems*, 2003.

[61] B. Le Saux, and N. Boujemaa, "Unsupervised Robust Clustering for Image Database Categorization," *Proc. International Conference on Pattern Recognition*, 2002.

[62] B. Li, K.-S. Goh, and E. Y. Chang, "Confidence-based Dynamic Ensemble for Image Annotation and Semantics Discovery," *ACM Multimedia*, 2003.

[63] J. Li, R. M. Gray, and R. A. Olshen, "Multiresolution Image Classification by Hierarchical Modeling with Two Dimensional Hidden Markov Models," *IEEE Trans. Information Theory*, 46(5):1826–1841, 2000.

[64] J. Li, D. Joshi, and J. Z. Wang, "Stochastic Modeling of Volume Images with a 3-D Hidden Markov Model," *Proc. IEEE International Conference on Image Processing*, 2004.

[65] J. Li and J. Z. Wang, "Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(9):1075–1088, 2003.

[66] J. Li and H.-H. Sun, "On Interactive Browsing of Large Images," *IEEE Trans. Multimedia*, 5(4):581–590, 2003.

[67] Y. Lu, C. Hu, X. Zhu, H.J. Zhang, and Q. Yang, "A Unified Framework for Semantics and Feature Based Relevance Feedback in Image Retrieval Systems," *Proc. ACM Multimedia*, 2000.

[68] J. Malik, S. Belongie, T. K. Leung, and J. Shi, "Contour and Texture Analysis for Image Segmentation," *International Journal of Computer Vision*, 43(1):7-27, 2001.

[69] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and Texture Descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, 11(6):703–715, 2001.

[70] K. Mikolajczyk and C. Schmid, "Scale and Affine Invariant Interest Point Detectors," *International Journal of Computer Vision*, 60(1):63–86, 2004.

[71] K. Mikolajczk and C. Schmid, "A performance evaluation of local descriptors," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[72] P. Mitra, C.A. Murthy, and S.K. Pal, "Unsupervised Feature Selection Using Feature Similarity," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(3):301–312, 2002.

[73] F. Monay and D. Gatica-Perez, "On Image Auto-Annotation with Latent Space Models," *Proc. ACM Multimedia*, 2003.

[74] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A Review of Content-Based Image Retrieval Systems in Medical Applications - Clinical Benefits and Future Directions," *International Journal of Medical Informatics*, 73(1):1–23, 2004.

[75] H. Muller, W. Muller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance evaluation in content-based image retrieval: Overview and proposals," *Pattern Recognition Letters*, 22(5):593-601, 2001.

[76] A. Natsev, R. Rastogi, and K. Shim, "WALRUS: A Similarity Retrieval Algorithm for Image Databases," *IEEE Trans. Knowledge and Data Engineering*, 16(3):301–316, 2004

[77] A. Natsev and J.R. Smith, "A Study of Image Retrieval by Anchoring," *Proc. IEEE International Conference on Multimedia and Expo*, 2002.

[78] K. Nakano and E. Takamichi, "An Image Retrieval System Using FPGAs," *Proc. Asia and South Pacific Design Automation Conference*, 2003.

[79] M. Nakazato, C. Dagli, and T.S. Huang, "Evaluating Group-based Relevance Feedback for Content-based Image retrieval," *Proc. IEEE International Conference on Image Processing*, 2003.

[80] T. H. Painter, J. Dozier, D. A. Roberts, R. E. Davis, and R. O. Green, "Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data," *Remote Sensing of Environment*, 85(1):64–77, 2003.

[81] E. G. M. Petrakis, A. Diplaros, and E. Milios, "Matching and Retrieval of Distorted and Occluded Shapes Using Dynamic Programming," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[82] M. Pi, M. K. Mandal, and A. Basu, "Image Retrieval Based on Histogram of Fractal Parameters," *IEEE Trans. Multimedia*, 7(4):597–605, 2005.

[83] K. Rodden, W. Basalaj, D. Sinclair, and K. Wood, "Does Organisation by Similarity Assist Image Browsing?," *Proc. ACM CHI*, 2001.

[84] K. Rodden and K. Wood, "How Do People Manage Their Digital Photographs?," *Proc. ACM CHI*, 2003.

[85] Y. Rui and T. S. Huang, "Optimizing Learning In Image Retrieval," *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 2000.

[86] Y. Rui, T.S. Huang, and S.-F. Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues," *Journal of Visual Communication and Image Representation*, 10(4):39-62, 1999.

[87] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval," *IEEE Trans. Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[88] M. Schroder, H. Rehrauer, K. Seidel, M. Datcu, "Interactive learning and probabilistic retrieval in remote sensing image archives," *IEEE Trans. Geoscience and Remote Sensing*, 38(5):2288–2298, 2000.

[89] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[90] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[91] Z. Su, H.-J. Zhang, S. Li, and S. Ma, "Relevance Feedback in Content-Based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressive Learning," *IEEE Trans. Image Processing*, 12(8):924–937, 2003.

[92] C. Theoharatos, N. A. Laskaris, G. Economou, and S. Fotopoulos, "A Generic Scheme for Color Image Retrieval Based on the Multivariate Wald-Wolfowitz Test," *IEEE Trans. Knowledge and Data Engineering*, 17(6):808–819, 2005.

[93] Q. Tian, N. Sebe, M. S. Lew, E. Loupias, and T. S. Huang, "Image retrieval using wavelet-based salient points," *Journal of Electronic Imaging*, 10(4):835–849, 2001.

[94] K. Tieu and P. Viola, "Boosting Image Retrieval," *Intl. Journal of Computer Vision*, 56(1/2), 17-36, 2004.

[95] S. Tong and E. Chang, "Support Vector Machine Active Learning for Image Retrieval," *ACM Multimedia*, 2001.

[96] R. S. Torres, C. G. Silva, C. B. Medeiros, and H. V. Rocha, "Visual Structures for Image Browsing," *Proc. ACM CIKM*, 2003.

[97] Z. Tu and S.-C. Zhu, "Image Segmentation by Data-Driven Markov Chain Monte Carlo," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(5):657–673, 2002.

[98] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image Classification for Content-Based Indexing," *IEEE Trans. Image Processing*, 10(1):117–130, 2001.

[99] N. Vasconcelos, "On the Efficient Evaluation of Probabilistic Similarity Functions for Image Retrieval," *IEEE Trans. Information Theory*, 50(7):1482–1496, 2004.

[100] N. Vasconcelos and A. Lippman, "Learning from User Feedback in Image Retrieval Systems," *Proc. Neural Information Processing Systems*, 1999.

[101] N. Vasconcelos and A. Lippman,"A Multiresolution Manifold Distance for Invariant Image Similarity," *IEEE Trans. Multimedia*, 7(1):127–142, 2005.

[102] N. Vasconcelos and A. Lippman, "A Probabilistic Architecture for Content-based Image Retrieval," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[103] J. Z. Wang, J. Li, R. M. Gray and G. Wiederhold, "Unsupervised Multiresolution Segmentation for Images with Low Depth of Field," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(1):85–90, 2001.

[104] J.Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(9), 947–963, 2001.

[105] Z. Wang, Z. Chi, and D. Feng, "Fuzzy integral for leaf image retrieval," *Proc. IEEE Intl. Conference on Fuzzy Systems*, 2002.

[106] E. Woodrow and W. Heinzelman, "SPIN-IT: A Data Centric Routing Protocol for Image Retrieval in Wireless Networks," *Proc. IEEE International Conference on Image Processing*, 2002.

[107] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, and V. Vapnik, "Feature Selection for SVMs," *Advances in Neural Information Processing Systems*, 2000.

[108] H. Wu, H. Lu, and S. Ma, "WillHunter: Interactive Image Retrieval with Multilevel Relevance Measurement," *Proc. IEEE International Conference on Pattern recognition*, 2004.

[109] P. Wu and B. S. Manjunath, "Adaptive Nearest Neighbor Search for Relevance Feedback in Large Image Databases," *Proc. ACM Multimedia*, 2001.

[110] Y. Wu, Q. Tian, and T. S. Huang, "Discriminant-EM Algorithm with Application to Image Retrieval," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[111] X. Xie, H. Liu, S. Goumaz, and W.-Y. Ma, "Learning User Interest for Image Browsing on Small-form-factor Devices," *Proc. ACM CHI*, 2005.

[112] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst, "Faceted Metadata for Image Search and Browsing," *Proc. ACM CHI*, 2003.

[113] S. X. Yu and J. Shi, "Segmentation Given Partial Grouping Constraints," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(2):173–183, 2004.

[114] Q. Zhang, S. A. Goldman, W. Yu, and J. E. Fritts, "Content-Based image retrieval using multiple-instance learning," *Proc. International Conference on Machine Learning*, 2002.

[115] Y. Zhang, M. Brady, and S. Smith, "Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm," *IEEE Trans. Medical Imaging*, 20(1):45–57, 2001.

[116] X. Zheng, D. Cai, X. He, W.-Y. Ma, X. Lin, "Locality Preserving Clustering for Image Database," *Proc. ACM Multimedia*, 2004.

[117] X. S. Zhou and T. S. Huang, "Unifying Keywords and Visual Contents in Image Retrieval," *IEEE Multimedia*, 9(2):23–33, 2002.

[118] X. S. Zhou and T. S. Huang, "Comparing Discriminating Transformations and SVM for Learning during Multimedia Retrieval," *Proc. ACM Multimedia*, 2001.

[119] X. S. Zhou and T. S. Huang, "Relevance Feedback in Image Retrieval: A Comprehensive Review," *Multimedia Systems*, 8:536-544, 2003.

[120] L. Zhu, A. Zhang, A. Rao, and R. Srihari, "Keyblock: An Approach for Content-based Image Retrieval," *Proc. ACM Multimedia*, 2000.