IMAGE RETRIEVAL USING CONTENT-BASED TECHNIQUE

B.C.G.Sharath Babu\textsuperscript{1}, M.Venkatesh Naik\textsuperscript{2}, S.Saravani\textsuperscript{3}

\textsuperscript{1}M.Tech Dept of CSE, Chiranjeevi Reddy Institute of Engineering & Technology, Affiliated to JNTUA, AP, India.
\textsuperscript{2}Assistant Professor, Dept of CSE, Chiranjeevi Reddy Institute of Engineering & Technology, Affiliated to JNTUA, AP, India.
\textsuperscript{3}Assistant Professor, Dept of CSE, Chiranjeevi Reddy Institute of Engineering & Technology, Affiliated to JNTUA, AP, India.

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.

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 Bibliotheca, 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, GoogleTM and Yahoo!r 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 highlevel 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. We searched for publications containing the phrases “Image Retrieval” using Google Scholar [3] 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.

II. 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 [2], 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 highdimensional 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 [3], 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. 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 [5]. 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 [1] which is fairly compact yet robust to a number of geometric transformations. A dynamic programming (DP) approach to shape matching has been proposed in [1]. 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]. 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 [6], 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 applicationspecific 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 ExpectationMaximization (EM) algorithm [1], and the spectral clustering approach has found some success in segmenting vertebral bodies from sagittal MR images [1]. Among other recent approaches proposed are segmentation based on the mean shift procedure [2], multi-resolution segmentation of low depth of field images [1], a Bayesian framework based segmentation involving the Markov chain Monte Carlo technique [1], and an EM algorithm based segmentation using a Gaussian mixture model [1], 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 [1]A discussion on the pros and cons of different types of color interest points used in image retrieval can be found in [3], while a comparative performance evaluation of the various proposed interest point detectors is reported in [1]. 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 [1]. Some of the other recent, more generic feature selection propositions involve boosting [6], evolutionary searching [5], Bayes classification error [1], and feature dependency/similarity measures [7]. A survey and performance comparison of some recent algorithms on the topic can be found in [3].

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 [9], here we briefly talk about some of the more recent approaches. A semantics-sensitive approach to content-based image retrieval has been proposed in [1]. A semantic categorization (e.g., graph - photograph, textured - nontextured) 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 [2]. Region based image retrieval has also been extended to incorporate spatial similarity using the Hausdorff distance on finite sized point sets [5], and to employ fuzziness to characterize segmented regions for the purpose of feature matching [1]. A framework for region-based image retrieval using region codebooks and learned region weights has been proposed in [4]. 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 colortexture segments called blobs, instead of image to image matching [2]. 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 [1]. 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 [4]. 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 [2]. A windowed search over location and scale has been shown more effective in object-based image retrieval than methods based on inaccurate segmentation [4]. 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 [2]. 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 [2]. Another wavelet-based retrieval method involving salient points has been proposed in [4]. Fractal block code based image histograms have been shown effective in retrieval on textured image databases [2]. The use of the MPEG-7 content descriptors to train self-organizing maps (SOM) for the purpose of image retrieval has been explored in [5]. Among other new approaches, anchoring-based image retrieval system has been proposed in [5]. 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 [3]. Clustering has been applied to image retrieval to help improve interface design, visualization, and result pre-processing [1]. A statistical approach involving the Wald-Wolfowitz test for comparing non-parametric multivariate distributions has been used for color image retrieval [2], representing images as sets of vectors in the RGB-space. Multiple-instance learning was introduced to the CBIR community in [1]. A number of probabilistic frameworks for image retrieval have been proposed in the last few years [3]. The idea in [1] 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 [9], 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 [1] 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 contentbased 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 [1]. An extension of multiple-instance learning has been shown effective for categorization of images into semantic classes [1]. Learning concepts from user’s feedback and within a dynamically changing image database using Gaussian mixture models is discussed in [2]. An approach to soft annotation, using Bayes Point machines, to give images a confidence level for each trained semantic label has been explored in [1]. 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 [4]. 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 [2].

III. 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 [1]. 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 [1] 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 [3]. Screen-shots can be seen in Fig. 2. Based on our experience with implementing CBIR systems on realworld 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 metadata 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 [5].

IV. 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.

REFERENCES