

A Survey on Content based Image Retrieval

Priya Singh
Computer Science & Engg. Department
S.I.T.M., Lucknow, India
priyasingh_8799@yahoo.com

Ravendra Ratan Singh
Computer Science & Engg. Department
S.I.T.M., Lucknow, India
ravendraratansingh@yahoo.co.in

Abstract:

Content-based Image Retrieval (CBIR), at present, poses to be a very lively discipline of research, expanding in its breadth. This thesis tries to bring out to the front the various challenges involved. The document describes the concepts of features, their selection, the aspect of databases etc. in order to develop a proper understanding of CBIR. Over the last decade, storage of non text-based data in databases has become an increasingly important trend in information management. Image in particular, has been gaining popularity as an alternative, and sometimes more viable, option for information storage. While this presents a wealth of information, it also creates a great problem in retrieving appropriate and relevant information during searching. This has resulted in an enormous growth of interest, and much active research, into the extraction of relevant information from non text-based databases. In particular, content-based image retrieval (CBIR) systems have been one of the most active areas of research.

1. Introduction:

Content-based image retrieval (CBIR), as we see it today, is any technology that in principle helps to organize digital picture archives by their visual content. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR. This characterization of CBIR as a field of study places it at a unique juncture within the scientific community. While we witness continued effort in solving the fundamental open problem of robust image understanding, we also see people from different fields, such as, computer vision, machine learning, information retrieval, human-computer interaction, database systems, Web and data mining, information theory, statistics, and psychology contributing and becoming part of the CBIR community [1]. Moreover, a lateral bridging of gaps between some of these research communities is being gradually brought about as a by-product of such contributions, the impact of which can potentially go beyond CBIR. Again, what we see today as a few cross-field publications may very well spring into new fields of study in the foreseeable future.

Amidst such marriages of fields, it is important to recognize the shortcomings of CBIR as a real-world technology. One problem with all current approaches is the reliance on visual similarity for judging semantic similarity, which may be problematic due to the semantic gap [2] between low-level

content and higher-level concepts. While this intrinsic difficulty in solving the core problem cannot be denied, we believe that the current state-of-the-art in CBIR holds enough promise and maturity to be useful for real-world applications if aggressive attempts are made.

For the purpose of completeness, and better readability for the uninitiated, we have introduced key contributions of the earlier years in Section 1. Image retrieval purely on the basis of textual metadata, Web link structures, or linguistic tags is excluded. The rest of this article is arranged as follows: For a CBIR systems to be useful in the real world, a number of issues need to be taken care of. Hence, the desiderate of real-world image retrieval systems, including various critical aspects of their design, are discussed in Section 2.2. Core research in CBIR has given birth to new problems, which we refer to here as CBIR offshoots. When distinct solutions to a problem as open-ended as CBIR are proposed, a natural question that arises is how to make a fair comparison among them.

2. Case Study:

GoogleTM and Yahoo!® are household names today primarily due to the benefits reaped through their use, despite the fact that robust text understanding is still an open problem.

Online photo-sharing has become extremely popular with [3], which hosts hundreds of millions of pictures with diverse content.

The video-sharing and distribution forum YouTube has also brought in a new revolution in multimedia usage. Of late, there is renewed interest in the media about potential real-world applications of CBIR and image analysis technologies, as evidenced by publications in Scientific American [4], Discovery News [5] and on [6].

We envision that image retrieval will enjoy a success story in the coming years. We also sense a paradigm shift in the goals of the next-generation CBIR researchers. The need of the hour is to establish how this technology can reach out to the common man in the way text retrieval techniques have. Methods for visual similarity, or even semantic similarity (if ever perfected), will remain techniques for building systems. What the average end-user can hope to gain from using such a system is a different question altogether.

Comprehensive surveys exist on the topic of CBIR [7, 8, 9], all of which deal primarily with work prior to the year 2000.

Surveys also exist on closely related topics such as relevance feedback [10], high-dimensional indexing of multimedia data [11], face recognition [10] (useful for face-based image retrieval), applications of CBIR to medicine, and applications to art and cultural imaging [12]. In our current survey, we restrict the discussion to image-related research only.

One of the reasons for writing this survey is that CBIR, as a field, has grown tremendously after the year 2000 in terms of the people involved and the papers published. Lateral growth has also occurred in terms of the associated research questions addressed, spanning various fields. To validate the hypothesis about growth in publications, we conducted a simple exercise. We searched for publications containing the phrases “Image Retrieval” using Google Scholar [13] and the digital libraries of ACM, IEEE, and Springer, within each year from 1995 to 2005. 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, the graph indicates similar growth patterns for both fields, although SVM has had faster growth. These trends indicate, given the implicit assumptions, a roughly exponential growth in interest in image retrieval and closely related topics. We also observe particularly strong growth over the last five years, spanning new techniques, support systems, and application domains. In this chapter, we comprehensively survey, analyze, and quantify current progress and future prospects of image retrieval. A possible organization of the various facets of image retrieval as a field is shown in Fig 2. Our article follows a similar structure. Note that the treatment is limited to progress mainly in the current decade, and only includes work that involves visual analysis in part or full.

2.1. Early Developments in Retrieval Techniques:

The years 1994–2000 can be thought of as the initial phase of research and development on image retrieval by content. The progress made during this phase was lucidly summarized at a high level in [2], which has had a clear influence on progress made in the current decade, and will undoubtedly continue to influence future work. Therefore, it is pertinent that we provide a brief summary of the ideas, influences, and trends of the early years (a large part of which originate in that survey) before describing the same for the new age.

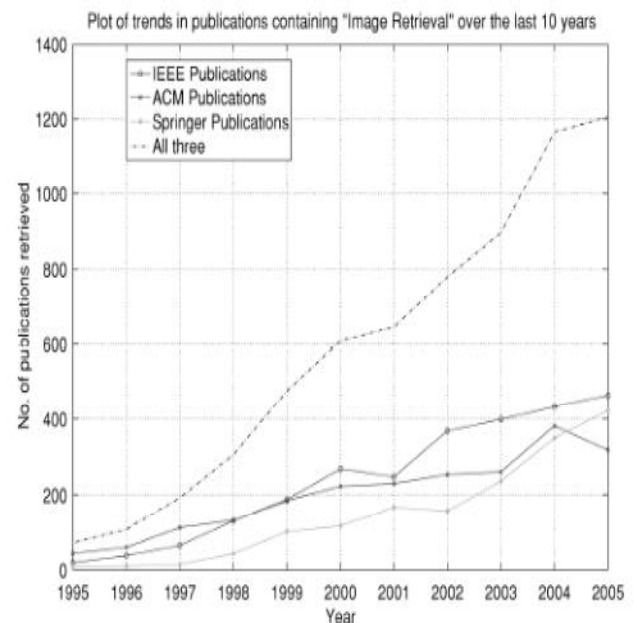
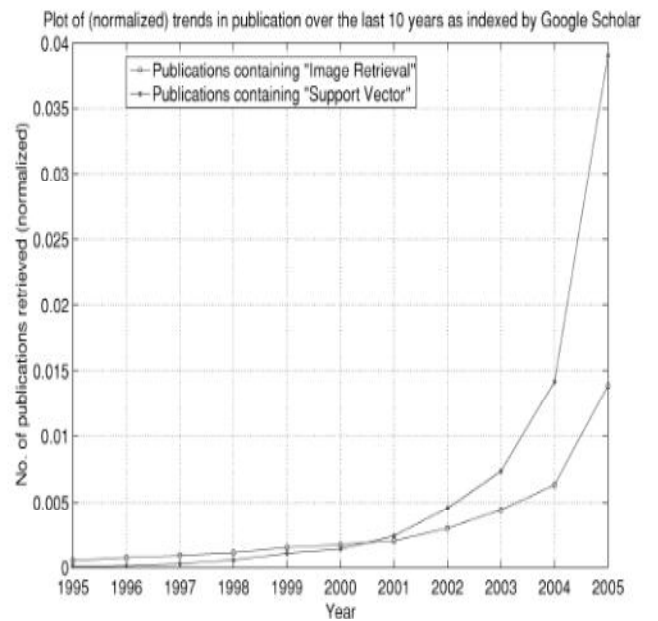


Fig. 1 A study of post-1995 publications in CBIR. Top: Normalized trends in publications containing phrases “image retrieval” and “support vector”. Bottom: Publisher, wise break-up of publication count on papers containing “image retrieval”.

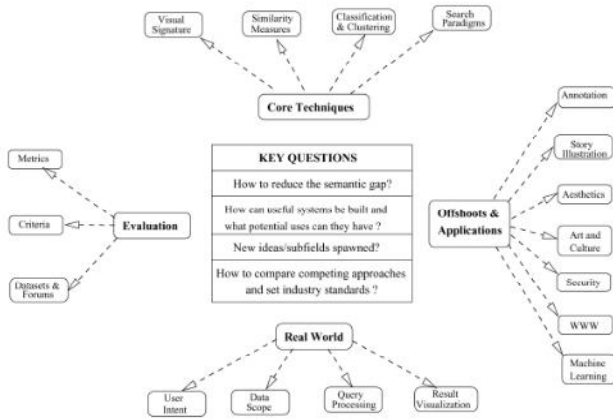


Fig. 2. Overview of many facets of image retrieval as a field of research.

The various gaps introduced here that define and motivate most of the related problems are given below:

- **Sensory:** The sensory gap is the gap between the object in the world and the information in a (computational) description derived from a recording of that scene.
- **Semantic:** The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data has for a user in a given situation.

While the former makes recognition from image content challenging due to limitations in recording, the latter brings in the issue of a user's interpretations of pictures and how it is inherently difficult for visual content to capture them. We continue briefly summarizing key contributions of the early years that deal with one or more of these gaps.

In [2], the domains for image search were classified as narrow and broad, and to-date this remains an extremely important distinction for the purpose of system design. As mentioned, narrow image domains usually have limited variability and better-defined visual characteristics (e.g., aviation-related pictures [14]), which makes content-based image search a tad bit easier to formulate. On the other hand, broad domains tend to have high variability and unpredictability for the same underlying semantic concepts (e.g., Web images), which makes generalization much more challenging. As recently noted in [15], narrow and broad domains pose a problem in image search evaluation as well, and appropriate modifications must be made to standard evaluation metrics for consistency.

The survey also lists three broad categories of image search: (1) search by association, where there is no clear intent at a picture, but instead the search proceeds by iteratively refined browsing; (2) aimed search, where a specific picture is sought; and (3) category search, where a single picture representative of a semantic class is sought, for example, to illustrate a paragraph of text, as introduced in [16]. Also discussed are different kinds of domain knowledge that can help reduce the sensory gap in image search. Notable among them are concepts of syntactic, perceptual, and topological similarity. The overall goal therefore remains to bridge the semantic and sensorial gaps using the available visual features of images and relevant domain knowledge to

support the varied search categories, ultimately to satiate the user. We discuss and extend some of these ideas from new perspectives in Section 2.

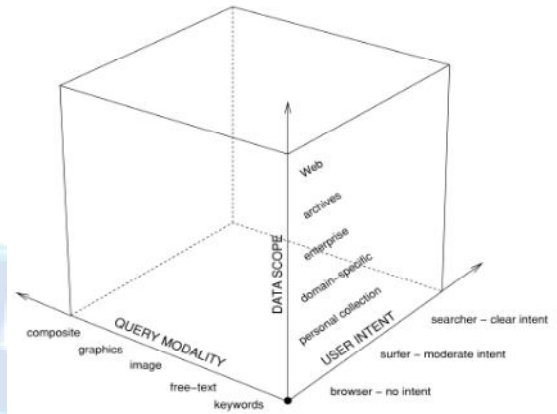
In the survey, extraction of visual content from images is split into two parts, namely image processing and feature construction. The question to ask here is what features to extract that will help perform meaningful retrieval. In this context, search has been described as a specification of minimal invariant conditions that model the user intent, geared at reducing the sensory gap due to accidental distortions, clutter, occlusion, etc. Key contributions in color, texture, and shape abstraction have then been discussed. Among the earliest use of color histograms for image indexing was that in [17]. Subsequently, feature extraction in systems such as QBIC [18], Pictoseek [19], and Visual SEEK [20] are notable. Innovations in color constancy, that is, the ability to perceive the same color amidst environmental changes, were made by taking specular reflection and shape into consideration [21]. In [22] color correlograms were proposed as enhancements to histograms, that take into consideration spatial distribution of colors as well. Gabor filters were successfully used for local shape extraction geared toward matching and retrieval in [23]. Daubechies' wavelet transforms were used to improve color layout feature extraction in the WBIIS system [24]. Viewpoint- and occlusion-invariant local features for image retrieval [25] received significant attention as a means to bridge the sensorial gap. Work on local patch-based salient features [26] found prominence in areas such as image retrieval and stereo matching. Perceptual grouping of images, important as it is for identifying objects in pictures, is also a very challenging problem. It has been categorized in the survey as strong/weak segmentation (data-driven grouping), partitioning (data-independent grouping, e.g., fixed-image blocks), and sign location (grouping based on a fixed template). Significant progress had been made in field of image segmentation, for example, where snake- and region growing ideas were combined within a principled framework, and [27], where spectral graph partitioning was employed for this purpose. From segments come shape and shape matching needs. In [28], elastic matching of images was successfully applied to sketch-based image retrieval. Image representation by multiscale contour models was studied in [29]. The use of graphs to represent spatial relationships between objects, specifically geared toward medical imaging, was explored in [30]. In [31], 2D-strings [32] were employed for characterizing spatial relationships among regions. A method for automatic feature selection was proposed in [33]. In [2], the topic of visual content description was concluded with a discussion on the advantages and problems of image segmentation, along with approaches that can avoid strong segmentation while still characterizing image structure well enough for image retrieval. In the current decade, many region-based methods for image retrieval have been proposed that do not depend on strong segmentation.

Once image features were extracted, the question remained as to how they could be indexed and matched against each other for retrieval. These methods essentially aimed to reduce the semantic gap as much as possible, sometimes

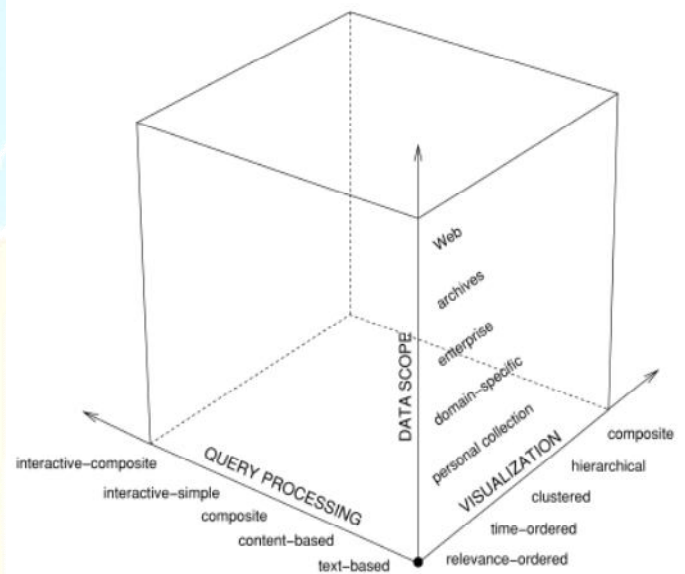
reducing the sensorial gap as well in the process. In [2], similarity measures were grouped as feature-based matching [17], object-silhouette-based matching [28], structural feature matching (i.e., hierarchically ordered sets of features, e.g., [34]), salient feature matching (e.g., [35]), matching at the semantic level (e.g., [36]), and learning-based approaches for similarity matching (e.g., [37] and [38]). Closely tied to the similarity measures are how they emulate the user needs, and, more practically, how they can be modified step-wise with feedback from the user. In this respect, a major advance made in the user interaction technology for image retrieval was relevance feedback (RF). Important early work that introduced RF into the image retrieval domain included [39], which was implemented in their MARS system [40]. Methods for visualization of image query results were explored, for example, in [18] and [41]. Content-based image retrieval systems that gained prominence in this era were, for example, IBM QBIC [18], VIRAGE [42], and NEC AMORE [43] in the commercial domain, and MIT Photobook [44], Columbia VisualSEEK and WebSEEK [20], UCSB NeTra [45], and Stanford WBIIS [24] in the academic domain. In [2], practical issues such as system implementation and architecture, as well as their limitations and how to overcome them, the user in the loop, intuitive result visualization, and system evaluation were discussed, and suggestions made. Innovations of the new age based on these suggestions and otherwise are covered extensively in our survey in Sections 2.

2.2. IMAGE RETRIEVAL IN THE REAL WORLD:

We devote this section to understanding image retrieval in the real world and discuss user expectations, system constraints and requirements, and the research effort to make image retrieval a reality in the not-too-distant future. Designing an omnipotent real-world image search engine capable of serving all categories of users requires understanding and characterizing user-system interaction and image search, from both user and system points-of-view. In Fig 3, shows one such dual characterization, and attempt to represent all known possibilities of interaction and search. From a user perspective, embarking on an image search, journey involves considering and making decisions on the following fronts: (1) clarity of the user about what he wants, (2) where he wants to search, and (3) the form in which the user has her query. In an alternative view from an image retrieval system perspective, a search translates to making arrangements as per the following factors: (1) how does the user wish the results to be presented, (2) where does the user desire to search, and (3) what is the nature of user input/interaction. These factors, with their respective possibilities, form our axes for Fig 3. In the proposed user and system spaces, real world image search instances can be considered as isolated points or point clouds, and search sessions can consist of trajectories while search engines can be thought of as surfaces. The intention of drawing cubes versus free 3D Cartesian spaces is to emphasize that the possibilities are indeed bounded by the size of the Web, the nature of user, and ways of user-system interaction. We believe that the proposed characterization will be useful for designing context-dependent search environments for real-world image retrieval systems.



(a) Visualixing image retrieval from a user perspective



(b) Visualixing image retrieval from a system perspective
Fig. 3. Our views of image retrieval from user and system perspectives.

3. Conclusion:

Many applications need to retrieve a set of images, from the image database, which are similar to the content of a given image. Such content-based image retrieval system usually will extract the features of the image and store them in a index file. The features to be considered include color, intensity and shape. Some researchers also include the spatial information for image analysis.

In this work, image retrieval methods based on color, shape and spatial analysis are investigated. We will design and implemented a prototype to retrieve a particular image from an image database. We will design an indexing methods based on different criteria. We introduce an integrated method that calculates the similarity value between two images. We then evaluate the performance and compare the characteristic of each image retrieval approach.

Reference:

- [1] wang, j. z., boujemaa, n., del bimbo, a., geman, d., hauptmann, a., and tesic, j. 2006. Diversity in multimedia information retrieval research. In Proceedings of the ACM Sigm International Workshop on Multimedia Information Retrieval (MIR) at the International Conference on Multimedia.
- [2] smeulders, a. w., worring, m., santini, s., gupta, a., and jain, r. 2000. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 12, 1349–1380.
- [3] Flickr. 2002. The flickr homepage. <http://www.flickr.com>.
- [4] Mirsky, S.. 2006. Computers get the picture. *Sci. Amer.* (Nov.).
- [5] Staedter, T. 2006. Digital pics 'read' by computer. http://dsc.discovery.com/news/2006/11/09/images_tec.html.
- [6] CNN. 2005. Computer decodes Mona Lisa's smile. <http://www.cnn.com/2005/TECH/12/16/mona.lisa.smile/index.html>.
- [7] Aigrain, P., Zhang, H., And Petkovic, D. 1996. Content-based representation and retrieval of visual media: A review of the state-of-the-art. *Multimed. Tools Appl.* 3, 3, 179–202.
- [8] rui, y., huang, t., and Chang, S.-F. 1999. Image retrieval: Current techniques, promising directions and open issues. *J. Visual Commun. Image Represent.* 10, 1, 39–62.
- [9] snoek, c. g. m. and worring, m. 2005. Multimodal video indexing: A review of the state-of-the-art. *Multimed. Tools Appl.* 25, 1, 5–35.
- [10] zhou, x. s. and huang, t. s. 2003. Relevance feedback in image retrieval: A comprehensive review. *Multimed. Syst.* 8, 536–544
- [11] bohm, c., berchtold, s., and keim, d. a. 2001. Searching in high-dimensional space index structures for improving the performance of multimedia databases. *ACM Comput. Surv.* 33, 3, 322–373.
- [12] chen, y., wang, j. z., and krovetz, R. 2005. Clue: Cluster-based retrieval of images by unsupervised learning. *IEEE Trans. Image Process.* 14, 8, 1187–1201.
- [13] google scholar. 2004. Google scholar homepage. <http://scholar.google.com>.
- [14] airliners.net. 2005. Airliners.net homepage. <http://www.airliners.net>.
- [15] huijsmans, d. p. and sebe, n. 2005. How to complete performance graphs in content-based image retrieval: Add generality and normalize scope. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 2, 245–251.
- [16] cox, i. J., miller, m. L., minka, t. P., papathomas, t. V., and yianilos, p. N. 2000. The bayesian image retrieval system, pichunter: theory, implementation, and psychophysical experiments. *Ieee trans. Image process.* 9, 1, 20–37.
- [17] swain, m. And ballard, b. 1991. Color indexing. *Int. J. Comput. Vision* 7, 1, 11–32.
- [18] flickner, m., sawhney, h., niblack, w., ashley, j., huang, q., dom, b., gorkani, m., hafner, j., lee, d., petkovic, d., steele, d., and yanker, p. 1995. Query by image and video content: the qbic system. *Ieee comput.* 28, 9, 23–32.
- [19] gevers, t. And smeulders, a. 2000. Pictoseek: combining color and shape invariant features for image retrieval. *Ieee trans. Image process.* 9, 1, 102–119.
- [20] smith, j. And chang, s.-f. 1997b. Visualseek: a fully automated content-based image query system. In proceedings of the acm international conference on multimedia.
- [21] finlayson, g. 1996. Color in perspective. *Ieee trans. Pattern anal. Mach. Intell.* 18, 10, 1034–1038.
- [22] huang, j., ravi kumar, s., mitra, m., zhu, w.-j., and zabih, r. 1999. Spatial color indexing and applications. *Int. J. Comput. Vision* 35, 3, 245–268.
- [23] manjunath, b. And ma, w.-y. 1996. Texture features for browsing and retrieval of image data. *Ieee trans. Pattern anal. Mach. Intell.* 18, 8, 837–842.
- [24] wang, j., wiederhold, g., firschein, o., and wei, s. 1998. Content-based image indexing and searching using daubechies' wavelets. *Int. J. Digital Librar.* 1, 4, 311–328.
- [25] schmid, c. And mohr, r. 1997. Local grayvalue invariants for image retrieval. *Ieee trans. Pattern anal. Mach. Intell.* 19, 5, 530–535.
- [26] tuytelaars, t. And van gool, l. 1999. Content-based image retrieval based on local affinity invariant regions. In proceedings of the international conference on visual information systems (visual).
- [27] shi, j. And malik, j. 2000. Normalized cuts and image segmentation. *Ieee trans. Pattern anal. Mach. Intell.* 22, 8, 888–905.
- [28] del bimbo, a. And pala, p. 1997. Visual image retrieval by elastic matching of user sketches. *Ieee trans. Pattern anal. Mach. Intell.* 19, 2, 121–132.
- [29] mokhtarian, f. 1995. Silhouette-based isolated object recognition through curvature scale space. *Ieee trans. Pattern anal. Mach. Intell.* 17, 5, 539–544.
- [30] petrakis, e. And faloutsos, a. 1997. Similarity searching in medical image databases. *Ieee trans. Knowl. Data eng.* 9, 3, 435–447.
- [31] smith, j. And chang, s.-f. 1997a. Integrated spatial and feature image query. *Ieee trans. Knowl. Data eng.* 9, 3, 435–447.
- [32] chang, s., shi, q., and yan, c. 1987. Iconic indexing by 2-d strings. *Ieee trans. Pattern anal. Mach. Intell.* 9, 3, 413–427.
- [33] swets, d. And weng, j. 1996. Using discriminant eigenfeatures for image retrieval. *Ieee trans. Pattern anal. Mach. Intell.* 18, 8, 831–836.
- [34] wilson, r. And hancock, e. 1997. Structural matching by discrete relaxation. *Ieee trans. Pattern anal. Mach. Intell.* 19, 6, 634–648.
- [34] wolfson, h. And rigoutsos, i. 1997. Geometric hashing: an overview. *Ieee trans. Comput. Sci. Eng.* 4, 4, 10–21.
- [36] fagin, r. 1997. Combining fuzzy information from multiple systems. In proceedings of the acm-sigactsigmod-sigart symposium on principles of database systems (pods).
- [37] wu, y., tian, q., and huang, t. S. 2000. Discriminant-em algorithm with application to image retrieval. In proceedings of the ieee conference on computer vision and pattern recognition (cvpr).
- [38] webe, m., welling, m., and perona, p. 2000. Unsupervised learning of models for recognition. In proceedings of the european conference on computer vision (eccv).
- [39] rui, y., huang, t. S., ortega, m., and mehrotra, s. 1998. Relevance feedback: a power tool in interactive content-



based image retrieval. Ieee trans. Circ. Syst. Video technol. 8, 5, 644– 655.

[40] rui, y., huang, t., and mehrotra, s. 1997. Content-based image retrieval with relevance feedback inmars. In proceedings of the ieee international conference on image processing (icip).

[41] chang, s., shi, q., and yan, c. 1987. Iconic indexing by 2-d strings. Ieee trans. Pattern anal. Mach. Intell. 9, 3, 413–427.

[42] gupta, a. And jain, r. 1997. Visual information retrieval. Commun. Acm 40, 5, 70–79.

[43] mukherjea, s., hirata, k., and hara, y. 1999. Amore: a world wide web image retrieval engine. In proceedings of the international world wide web conference (www).

[44] pentland, a., picard, r., and sclaroff, s. 1994. Photobook: tools for content-based manipulation of image databases. In proceedings of the conference on storage and retrieval for image and video database ii, spie, san jose, ca.

[45] ma, w. And manjunath, b. 1997. Netra: a toolbox for navigating large image databases. In proceedings of the ieee international conference on image processing (icip).