COMPUTER VISION AND INFORMATION TECHNOLOGY

Subhasis Chaudhuri and **Amit Prabhudesai**
Department of Electrical Engineering, Indian Institute of Technology, Bombay, Mumbai - 400076, India

Keywords: Content-based image retrieval, automatic illumination correction, image super-resolution, photo-albumation, virtual-view generation.

Contents

Summary

This chapter presents an overview of techniques to interpret, enhance, and augment digital image and/or video information. First, it provides an insight into methods to index an image database, and enables the user to extract the information he/she wants from a huge image database. This is the role of a content-based image retrieval system. The next section explains a method for scene enhancement in cases where the original scene may be poorly illuminated. An application where this method may be used is presented. The chapter also discusses a unique approach to enhance the spatial resolution of an image. The next section illustrates a novel approach to automatically extract frames from a video sequence. This can be used for automatic generation of a high resolution photo album. Finally, the concluding section looks at a method to generate 'virtual views'- views captured with a different camera parameter setting or from a different camera position, given a sequence of views.

1. Introduction

There has been a massive boost in information and communication technology since the 1990s. This has been facilitated by recent developments in five key technology areas- processors, storage, communication, sensor and display. Large-scale deployment of wireless technology has allowed us to enjoy multimedia applications or services anytime or anywhere. But having the system or the services alone is not enough. One must have the multimedia contents so that a specific service can by provided to the user. Further, having created such contents, one realizes that we should also have an efficient content management system to handle such a large amount of data. The purpose of this chapter is to illustrate recent progresses in content creation and management. Due to space restrictions, we address only a limited number of issues here. Further we concentrate only on the image and video aspects of multimedia.

One will argue that creation of content is an art and we do agree. However, we show that technology can help an artist in creating content as per his requirements. For example, an artist may want to capture a picture of a celebrity or a building from a specific vantage point or with a specific light source position that he or she does not have access to. The technology can aid him or her in such a case by allowing the artist to create virtual views. If the content creator is an amateur, then the technology can offer an even bigger helping hand.

The technology also helps the end-users. If one is overwhelmed by the sheer volume of digital data that he or she has to browse to search for some particular content, technology can help him or her to retrieve it. And that too at only a click of a button! If the user is more interested in seeing an image in greater detail, we can provide this by improving its spatial resolution. Also suppose the user wants to save only some important frames from a long duration video, again the technology comes to his or her aid. In the next few sections, we explain in very simple terms how the above tasks can be achieved. Bibliographical references are provided to guide the interested reader to more detailed material about the topics covered in these sections.

2. Content Based Image Retrieval

The Internet has revolutionized the way we access information. Where earlier one had to pore over library catalogues and index pages of a text to search for relevant information, today almost anything that we might want is a few mouse-clicks away. It is the Internet search engines like Google, Yahoo or MSN to name a few, which have made it possible to access almost any information, from a recipe for an apple pie to the most advanced research papers on any subject, stored in the remotest corners of the globe in a jiffy. The main content being searched in these cases is primarily text. For instance, the keywords that a user types in Google search are used by the search engine to return documents (ranked in order of decreasing relevance) that contain the keywords that the user might be interested in. In this section we look at a different form of querying. What if the content we were interested in was not text, but say, an image? Google's image search does cater to this need, albeit in a restricted fashion. The results returned by the system are indeed images, but the query is still textual. There may be situations when one's information needs might be better expressed by an image rather than by text. After all, a picture speaks better than a thousand words! It is this situation that we look at in the present section.

The last decade of the twentieth century witnessed rapid advances in the Internet and digital image sensor technologies. This among other things contributed to the vast amount of images produced by scientific, medical, educational, industrial and other applications. This led to huge databases of images whose content spans a wide range. A content based image retrieval (CBIR) system is required to efficiently use information from these image repositories. Such a system helps users (even those unfamiliar with the database) retrieve relevant images based on their content. Application areas in which CBIR is a principal activity are numerous and diverse. Only a few of these are listed below:

- art galleries and museum management
- architectural and engineering design
- geographic information systems
- picture archiving and communication systems,
- law enforcement and criminal investigations, and
- mining a database of medical images

At a first glance, content based querying appears deceptively simple because we humans are so good at it. The moment we see a picture, we subconsciously extract the relevant semantic information from it. Extracting semantic information is as yet an unsolved problem in computer vision. Computers, however, are much better than humans at measuring properties and retaining these in long-term memory. Thus all CBIR systems essentially compute and store certain low-level features of the images in the database and compare these with the corresponding features of the query image, to find possible matches. The problem with this is that the semantic interpretation of an image for a particular user might not be the same as the information represented by the low-level features. This lack of coincidence is called the semantic gap, and poses a major challenge in image retrieval.

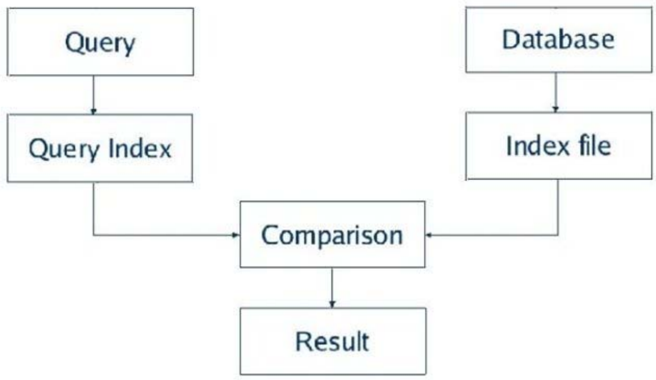

Figure 1. Block diagram of a content-based image retrieval system.

Basics of a CBIR system

A content-based image retrieval system uses visual contents of an image such as *color*, *shape*, *texture*, and *spatial layout* to represent and index the image. Figure 1 shows a block- diagram of a typical content-based image retrieval system. The visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. To retrieve images, users provide the retrieval system with example images or sketched figures. The system then changes these examples into its internal representation of feature vectors. The similarities /distances between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme. The indexing scheme provides an efficient way to search the image database. Recent retrieval systems have incorporated users' relevance feedback to modify the retrieval process in order to generate perceptually and semantically more meaningful techniques for CBIR.

2.1. Image Content Descriptors

This section introduces some widely used techniques for extracting the visual content of images.

2.1.1. Color

Color is the most extensively used visual content for image retrieval. It is used because of the inherent simplicity and intuitiveness of the color feature. Often it is used as a first pass to eliminate images which are *not* likely to be relevant. For example, to retrieve images of sunsets from an image database, the system can eliminate images that do not have a predominance of red hues.

A digitized image is represented in a certain *color space*. Simply put, a color space provides us with a set of co-ordinate axes in which to represent a given point (pixel) in an image. A choice of an appropriate color space is necessary before selecting an appropriate color description. Commonly used color spaces for image retrieval include *RGB, CIE L*a*b*, CIE L*u*v*, HSV* and *opponent* color spaces. One of the desirable characteristics of an appropriate color space for image retrieval is its uniformity. Simply put, it means that the measured proximity among the colors must be directly related to the psychological similarity among them.

Color descriptors: Some of the most commonly used descriptors include color moments, color histogram, color coherence vector, and color correlogram.

- *Color histograms*: The color histogram attempts to capture the global color distribution of the image. It is easy to compute, robust to translation and rotation, and changes only slowly with scale, occlusion and viewing angle. However, with large number of images in the database, histogram comparison will saturate the discrimination. In addition, the global color histogram does not capture the spatial information. A local color histogram solves this problem to some extent. The trade-off is between increased computation time and precision in the retrieved results.
- *Color coherence vector*: The color coherence vector is a different way of incorporating spatial information into the color histogram. Basically, each histogram bin is partitioned into two types, i.e., coherent and incoherent. It has been shown that the color coherence vector provides better retrieval results than the color histogram, especially for those images which have mostly uniform color or mostly texture regions.
- *Color correlogram*: The color correlogram was proposed to characterize not only the color distributions of pixels, but also the spatial correlation of pairs of colors. A color correlogram is a table indexed by color pairs, where the k-th entry for (i, j) specifies the probability of finding a pixel of color j at a distance k from a pixel of color i in the image. Compared to the color histogram and the color coherence vector, the color autocorrelogram provides the best retrieval results, but is also the most computationally intensive due to its high dimensionality.

2.1.2. Texture

The problem with using only color for image retrieval is that vastly different scenes may have the same color. In the simple example of a sunset scene cited above, it may be noted that using only color to retrieve images would also return images of, say apples, which have similar color content. Thus, color alone is not very effective for image retrieval. The human visual system also uses other cues like the texture of an object and/or scene to recognize a particular image. For example, it is the texture cue which helps us to differentiate between a lawn and a tree, although both have similar colors. We note that texture conveys information about a region, or a neighborhood of a particular pixel, rather than only the pixel. It is this additional information that provides for better image retrieval. Various texture representations have been investigated in computer vision and pattern recognition literature. A few of these are listed below.

- *Tamura and Wold features*: The human perception of texture has been proved to be based on characteristics like coarseness, contrast, directionality, regularity (or the lack of it!), and roughness. These characteristics form the guiding principles in the Tamura features and Wold features, which have been widely used in image retrieval.
- *Gabor filter features:* The Gabor filter has been widely used to extract image features, especially texture features. Gabor filters are defined by harmonic functions modulated by a Gaussian function. Gabor filters bear some similarities to Fourier filters, but (due to the Gaussian damping terms) are limited to certain frequency bands. Gabor filters have often been used as an orientation and scale tunable edge and line (bar) detector.
- *Wavelet transform feature:* Similar to the Gabor filtering, the wavelet transform provides a multi-resolution approach to texture analysis and classification. Wavelet transforms decompose a signal with a family of basis functions obtained through translation and dilation of a mother wavelet.

2.1.3. Shape

Often images can be segmented distinctly into foreground and background regions. The foreground regions may consist of one or more *objects*. Here the word 'objects' must be taken in a broad sense to imply a contiguous region representing some known entity, such as, for example, a human being, an automobile, a building, etc. These objects may be used in comparing images and in their retrieval. For example, if we are searching for images of wild animals, then the shape information may be useful in retrieving images. Another practical example where the shape information may be cleverly used is an application requiring retrieval of logos (of a particular company or brand). These objects may be described by using their shape properties. A good shape representation feature for an object should be invariant to translation, rotation and scaling. Classical shape representation uses a set of *moment invariants*. If the object R is represented as a binary image, then the central moments of order $p+q$ for the shape of object R are defined as

$$\mu_{p,q} = \sum_{(x,y)\in R} (x-x_c)^p (y-y_c)^q$$

where (x_c, y_c) is the centroid of the object. This central moment may be normalized to be scale invariant. Based on these moments, a set of moments invariant to translation, rotation, and scale may be derived.

Boundary based methods are also used to represent shape information. The contour of a 2D object can be represented as a closed sequence of successive pixels (x_s, y_s), where $0 \le s \le N-1$ and N is the total number of pixels on the boundary. This representation of an object is used to calculate the *turning function* or the *turning angle* $\theta(s)$, which measures the angle of the counter-clockwise tangents as a function of arc-length s

according to a reference point on the contour. Instead of boundary, one can also use geometric matching of spatial patterns for computational benefits.

Spatial information: Regions or objects with similar color and texture properties can be easily distinguished by imposing spatial constraints. For instance, regions of blue sky and ocean may have similar color histograms, but their spatial locations in images are different. Therefore, the spatial location of regions (or objects) or the spatial relationship between multiple regions (or objects) in an image is very useful for searching images. 2D strings and its variants, spatial quad-tree and symbolic image are typically used for spatial information representation.

The primary image content descriptors are as noted above. Often though, a single descriptor is insufficient to obtain acceptable results. Hence, most state-of-the-art systems use a combination of two or more of the descriptors. For example, a typical CBIR system may use a feature vector comprised of the color moments and the Gabor coefficients. The image content representable by the combined use of both color and texture is much more effective than using any single feature alone.

2.2. Similarity/Distance Measures

Instead of exact matching, CBIR calculates visual similarity between a query image and images in a database. Accordingly, the retrieval result is not a single query image but a list of images ranked by their similarities with the query image. Different similarity/distance measures will affect the retrieval performance of an image retrieval system significantly. A very commonly used distance measure is the *quadratic-form* distance. This is simply the sum of squared differences for each element in the image descriptor vectors of the query image and one of the database images. This is a simple distance measure that assumes that each element of the vector is of equal importance in calculating the distance. In case this is not the case, a different distance measure, called the *Mahalanobis distance* is used. The Mahalanobis distance makes provision for assigning weights to the different elements in the image descriptor. Some other descriptors have also been used in image retrieval systems. These include, among others,

- Minkowski-form distance
- Histogram intersection
- Kullback-Leibler divergence
- Earth Mover's distance and
- Hausdorff distance.

The target application plays a role in deciding which of these distance measures is most appropriate. Very often applications also provide results using different distance measures separately. The user then selects the most relevant results.

2.3. Indexing Schemes

Another important issue in CBIR is effective indexing and fast searching of images based on visual features. Because the feature vectors of images tend to have high dimensionality and therefore are not well suited to traditional indexing structures, dimension reduction is usually used before setting up an efficient indexing scheme. For this purpose, techniques like principal component analysis (PCA), Karhunen-Loeve transform, and neural networks have been used.

After dimension reduction, the multi-dimensional data are indexed. A number of approaches have been proposed for this purpose, including R-tree, linear quad-trees, K-d-B tree and grid files.

2.4. Some Interesting Applications

This section looks at some interesting applications of the concepts just discussed. We look at diverse applications that will illustrate few of the methods that are used in current state-of-the-art systems. The first application involves searching for logos. This is followed by a personal image collection browser, wherein a person may need to retrieve some specific image(s) from his/her personal image collection. The third application could be useful to architects/designers to help them retrieve images of buildings or architectural monuments.

Figure 2. Example showing retrieval of logos.

-
-
-

Bibliography

Barnard K., Cardei V. and Funt B. (2002). A Comparison of Computational Color constancy algorithms- Part I: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing.* 11(9): 972-984, September 2002. [This and the following paper explain various methods of achieving color constancy.]

Barnard K., Martin L., Coath A., and Funt B. (2002). A Comparison of Computational Color constancy algorithms- Part II: Methodology and experiments with image data. *IEEE Transactions on Image Processing.* 11(9): 985-996, September 2002.

Bishop C.M., Blake A., and Marthi B. (2003). Super-resolution enhancement of video. In *Int. Conf. on Artificial Intelligence and Statistics*, Key West, Florida. [A paper on how to do super-resolution.]

Burrus. C., Gopinath R.A., and Guo. H. (1998). *Introduction to Wavelets and Wavelet Transforms.*

Prentice Hall, New Jersey. [The concept of wavelet is explained.]

Chang S.K., Shi Q.Y. and Yan C.Y. (1987). Iconic indexing by 2-D strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(3): 413-428, May 1987. [The issue of indexing in databases is discussed here.]

Chang T. and Kuo C.C.J. (1993). Texture analysis using tree-structured wavelet transform. *IEEE Transactions on Image Processing*, 2(4):429-441, October 1993. [This is about the usefulness of wavelet transform in analyzing image texture.]

Daubechies I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia, Pennsylvania. [It provides a very good description of the wavelet concept.]

Elad M. and Feuer A. (1999). Super-resolution on an image sequence: an Adaptive Filtering approach. *IEEE Transactions on Image Processing*, 8(3):387-395, March 1999. [This is about how an image can be super-resolved using an adaptive filter.]

Gudivada V.N. and Raghavan V.V. (1995). Design and Evaluation of Algorithms for Image Retrieval by Spatial Similarity. *ACM Transactions on Information Systems*, 13(2):115-144, April 1995. [This paper focuses on the usage of spatial similarity in contents while searching for images.]

Hartley H. and Zisserman A. (2003). *Multiple View Geometry in Computer Vision.* Cambridge University Press. 2nd edition. [This is a graduate level text book on applications of projective geometry in computer vision.]

Ioka M. (1989). A method for defining the Similarity of images on the basis of Color information. *Technical Report RT-0030*, November 1989. [This report discusses retrievability of images based on the color feature.]

J. Huang. (1997). Image Indexing using Color Correlogram. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 762-768, Puerto Rico, June 1997. [The usage of color correlogram in CBIR is discussed in this paper.]

Jain A.K. and Farroknia F. (1991). Unsupervised Texture Segmentation using Gabor Filters. *Pattern Recognition*, 24(12): 1167-1186. [Applicability of Gabor filters in texture image segmentation has been studied here.]

Joshi M.V. and Chaudhuri S. (2003). A learning-based method for Image Super-resolution from zoomed observations. In *Proc. 5th Int. Conf. on Advances in Pattern Recognition*, pp. 179-182, Indian Statistical Institute, Kolkata, India. [This paper shows that images of a scene captured with a variable amount of zooming can be used for image super-resolution.]

Liu. F. and Picard R.W. (1996). Periodicity, directionality and randomness: Wold Features for Image Modeling and Retrieval. *IEEE Transactions on Pattern Analysis and Machine Learning*, 18(7), July 1996. [This paper discusses various types of features that can be computed from an image and their applicability in CBIR.]

Nievergelt J., Hinterberger H., and Sevcik K.C. (1984). The grid file: an adaptable symmetric multi-key file structure. *ACM Transactions on Database Systems.* Vol.9, No. 1., pp. 38-71, March 1984. [This paper is useful to learn about indexing of image databases.]

Pass G. and Zabith R. (1996). Histogram Refinement for Content-based Image Retrieval. *IEEE Workshop on Applications of Computer Vision*, pp. 96-102. [This paper is about histogram and its use in CBIR.]

Robinson J.T. (1981). The k-d-B tree: a search structure for large multidimensional dynamic indexes. In *Proc. Of SIGMOD Conf*, Ann Arbor, April 1981. [This provides a searching strategy for indexed files.]

S. Baker and T. Kanade. (2002). Limits on Super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9): 1167-1183, September 2002. [This paper discusses super-resolvability of a scene from its views.]

Samet H. (1984). The quad-tree and related hierarchical data structures. *ACM Computing Surveys*, 16(2): 187-260. [This provides a good survey on data structures.]

Sivic J. and Zisserman A. (2003). Video Google: A text retrieval approach to object matching in videos. In *Proc. International Conf. on Computer Vision,* Nice, France. [It provides a new technique for searching image databases.]

Tamura H., Mori S. and Yamawaki T. (1978). Texture features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics.* pp. 460-472, June 1978. [It provides a description for various types of textural features in images.]

Biographical Sketches

Dr. Subhasis Chaudhuri received his B. Tech degree in Electronics and Electrical Communication Engineering from the Indian Institute of Technology, Kharagpur in 1985. He received his M.S. and Ph.D. degrees, both in Electrical Engineering, respectively, from the University of Calgary, Canada and the University of California, San Diego. He joined IIT Bombay in 1990 as an Assistant Professor and is currently serving as the Professor and Head of the Electrical Engineering Department.

He has also served as a visiting professor at the University of Erlangen-Nuremberg, Germany and the University of Paris XI. He is a fellow of the Alexander von Humboldt Foundation, Germany, the Indian National Academy of Engineering, the National Academy of Sciences, and the Indian Academy of Sciences. He is the recipient of the Vikram Sarabhai Research Award for the year 2001, and the *Swarnajayanti* Fellowship in 2003. He received the S.S. Bhatnagar Prize in Engineering Sciences for the year 2004.

He is the co-author of the following books: ‘Depth from defocus: a real aperture imaging approach’, and ‘Motion-free Super-resolution’, both published by Springer, NY. He has also edited a book on ‘Super-resolution Imaging’ published by Kluwer Academic in 2001. He is currently in the editorial board of the following journals: IEEE Transactions on Pattern Analysis and Machine Intelligence, International Journal of Computer Vision, IET Proceedings on Computer Vision, and *Sadhana*. His research interests include image processing, computer vision, and multimedia.

Amit Prabhudesai received his B.E. degree in Electronics Engineering from the University of Bombay in 2004, and the M. Tech degree in Electrical Engineering in 2006. He is currently working as a Research Engineer in Siemens Corporate Technology, Bangalore. His research interests include image processing and computer vision.