A Survey of Shape Descriptors for Digital Image Processing

CS PILLAI

RESEARCH STUDENT, VINAYAKA MISSON'S UNIVERSITY,SALEM,TAMILNADU.INDIA.pillai.cs5@gmail.com

Abstract

Automatic image retrieval is an important research problem considering its usage and ever increasing volume of image data existed in different kinds of databases both on the web as well as in the network computing system. Image retrieval is normally being done using label attachment to each of the image. This may be either a text document generated manually for annotation or feature map extracted from the image automatically. Former methods of manual labeling become irrelevant due to continuous increasing in size of image databases. The image retrieval based on feature extracted from inherent image characteristics automatically is known as content based image retrieval (CBIR), and is the current practice for image retrieval.

The basic feature of CBIR is used to return a group of images from a very large image database based on the query image given to the system. The accuracy of the system largely depends upon the quality of visual features used to represent an image information which can capture the overall visual impression of it. To compare the similarity of images from the database with query image, a different kind of similarity measures may be used.

Keywords: CBIR-Labeling-Query Image-Visual Impression-Feature Extraction- Image Retrieval

1. Introduction

Generally, low level features such as color, texture, shape, corner, etc., are used to represent the approximate perceptual representation of an image, using which similarity and dissimilarity of the images are computed. But it is found that the perceptual representation of an image in terms of low level features fails to capture entire semantic information of an image and it is often difficult to model accurately. The results obtained with these features lower the accuracy of a CBIR system than it is expected.

Moving Picture Expert Group (MPEG)-7 is an ISO/IEC standard developed by MPEG to facilitate effective uses of audio, visual (color, texture, shape, etc.) and motion picture description to address multimedia retrieval. It was formally named as multimedia content description interface. It is a standard for the multimedia content data that supports interpretation of the information, which can be passed onto, or accessed by a device or a computer code. MPEG-7 is not targeted at any one application in particular; rather it supports a range of applications as possible.

MPEG-7 descriptor feature based retrieval is very extensive and can capture very closely low level visual description through number of descriptors. It induces low level feature extraction algorithms using color, texture, motion, and shape, facilitated for image and video retrieval and benchmarking of newly proposed schemes. MPEG-7 provides many descriptors for retrieval.

- CONTOUR BASED SHAPE DESCRIPTOR
- REGION BASED SHAPE DESCRIPTOR

1.1 Curvature Scale Space Descriptor (CSSD)

In this section, CSSD is described in detail. The computation of CSSD is given in algorithm forms to make it more convenient for implementation.

Computing Curvature Scale Space Descriptor: Basically, the CSS method treats shape boundary as a 1D signal, and analyzes this 1D signal in scale space. By examining

zero crossings of curvature at different scales, the concavities/convexities of shape contour are found. These concavities/convexities are useful for shape description because they represent the perceptual features of shape contour.

The first step of the process is the same as that in computing FD – the output is the boundary coordinates $(x(t)y(t)), t = 0, 1, 2, . . . , N-1$. The second step is scale normalization, which samples the entire shape boundary into a fixed number

of points so that shapes with a different number of boundary points can be matched. The other two main steps in the process are the *CSS contour map* computation and *CSS peaks* extraction. The CSS contour map is a multi-scale organization of the inflection points (or curvature zero-crossing points). To calculate the CSS contour map, curvature is first derived from shape boundary points ($x(t)y(t)$), $t = 0, 1, 2, \ldots, N-1$:

$$k(t)=(x(t)\quad y(t)-x(t)y(t))/(x^2(t)+y^2(t))^{3/2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

where $x^{\cdot}(t)$, $y^{\cdot}(t)$ and $x^{\cdot\cdot}(t)$, $y^{\cdot\cdot}(t)$ are the first and the second derivatives at location t, respectively. Curvature *zero-cross points* are then located in the shape boundary. The shape is then evolved into the next scale by applying Gaussian smoothing:

$$X'(t)=x(t)\quad *g(t,\alpha)\qquad y'(t)=y(t)\ *g(t,\alpha) \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

where $*$ means convolution, and $g(t, \sigma)$ is *Gaussian function*. As σ increases, the evolving shape becomes smoother. New curvature zero-crossing points are located at each scale. This process continues until no curvature zero-crossing points are found. The evolution process is demonstrated in Fig. 1.

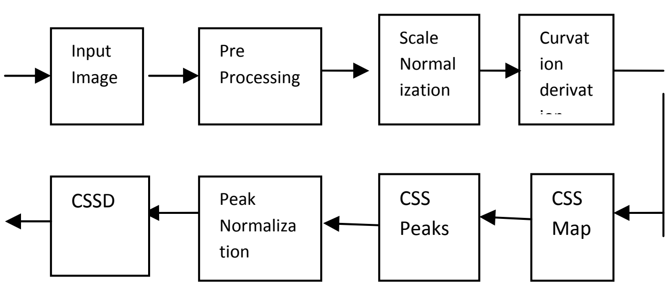

Fig1:-Block diagram of computing CSSD

1.2 Zernike Moment Descriptor

The use of orthogonal moments to recover the image from moments based on the theory of orthogonal polynomials, and introduced Zernike moments, which allow independent moment invariants to be constructed to an arbitrarily high order. The complex Zernike moments are derived from Zernike polynomials:

$$Vnm(x, y) = Vnm(\rho \cos \theta, \rho \sin \theta) = Rnm(\rho) \exp(jm\theta) \quad (5)$$

Where

$$R_{nm}(\rho)= \sum_{(s=0)}^{((n-|m|)/2)} (-1)^s (n-s)!/(s!((n+|m|)/2-s)!((n-|m|)/2)!)\, \rho^{n-2s}$$

where ρ is the radius from (x, y) to the shape centroid, θ is the angle between ρ and the x-axis, and n and m are integers and subject to $n - |m| = even$, $|m| \le n$. Zernike polynomials are a complete set of complex-valued function orthogonals over the unit disk, i.e., $x2+y2 = 1$. The complex Zernike moments of order n with repetition m are then defined as:

$$A_{nm}=\frac{N+1}{\pi}\sum_x\sum_y f(x,y)v_{nm}(x,y) \quad ,x^2+y^2\le 1$$

where * means complex conjugate. Due to the constraint of $n-|m| = even$ and $m < n$, there are $n/2$ repetitions of moments in each order n.

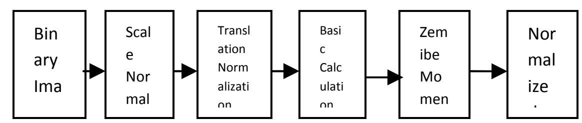

Fig2:- Block diagram of computing ZMD

Since Zernike basis functions take the unit disk as their domain, this disk must be specified before moments can be calculated. In our implementation, all shapes are normalized into a unit circle of fixed radius of 64 pixels. The unit disk is then centered on the *center of mass* of the shape. This makes the moments obtained both scale and translation invariant. Rotation invariance is achieved using only the magnitudes of the moments.

The magnitudes are then normalized into [0, 1] by dividing them by the mass of the shape. The similarity between two shapes indexed with Zernike moments descriptors is determined by the city block distance between the two Zernike moments vectors. A block diagram of the whole process of computing ZMD is shown in Fig. 11.

The theory of Zernike moments is similar to that of Fourier transform, to expand a signal into a series of orthogonal basis. However, the computation of a Zernike moments descriptor does not need to know boundary information, making it suitable for more complex shape representation. Like the Fourier descriptor, Zernike moment invariants can be constructed to arbitrary order, thus overcoming the drawback of geometric moment in which higher order moment invariants are difficult to construct. The precision of shape representation depends upon the number of moments truncated from the expansion. For efficient retrieval, the first 36 moments of up to order 10 are used in our implementation.

1.3 Short Time Fourier Transform

Short-time Fourier transform (STFT) of the data, the arrival time of the first peak is shown to be faster and more distinctive than that without STFT if the observation frequency is properly selected and investigated the characteristic features of the signature by extracting the first and second peaks of the received signal in time domain. The arrival time and amplitude of the first peak and the time delay of the second peak have been mainly used in the analysis of measured radar data because of the simplicity and accuracy of these parameters. The most important parameter is the time-of-peak (TOP), which denotes the arrival time of the first peak.

1.4 Wavelet Based Gradient Point

Data representation and content description are two basic components required by the management of any image database. A wavelet-based system, called the WaveGuide, which integrates these two components in a unified framework, is proposed in this work. In the WaveGuide system, images are compressed with the state-of-the-art wavelet coding technique and indexed with color, texture, and object shape descriptors generated in the wavelet domain during the encoding process. All the content descriptors are extracted by machines automatically with a low computational complexity and stored with a low memory space. Extensive experiments are performed to demonstrate the performance of the new approach.

Due to the tremendous growth of multimedia information, effective management of multimedia archiving and storage systems becomes more important and challenging.

For example, a remote sensing satellite, which generates seven band images including three visible and four infrared spectrum regions, produces around 5000 images per week. Each single spectral image, which corresponds to a 170 km 185 km of the earth region, requires 200 Mega bytes of storage. It is estimated that the amount of data originated from satellite systems will reach a terabyte per day. To store, index, and retrieve such a huge amount of data is a very challenging task. Generally speaking, data representation and content description are two basic components required by the management of any multimedia database. As far as the image database is concerned, the former is concerned with image storage while the latter is related to image indexing and retrieval.

Current commercial systems support image indexing based on the use of keywords or text phrases associated with images. The keyword is a high-level tool of content description, and has been successfully applied to textual databases.

1.5 Sift/Surf:(Scale Invariant Feature Transform/Speeded Up Robust Features)

As datasets grow increasingly large in content-based image and video retrieval, computational efficiency of concept classification is important. This paper reviews techniques to accelerate concept classification, where we show the trade-off between computational efficiency and accuracy. As a basis, we use the Bag-of-Words algorithm that in the 2008 benchmarks of TRECVID and PASCAL lead to the best performance scores. We divide the evaluation in three steps:

1) Descriptor Extraction, where we evaluate SIFT, SURF, DAISY, and Semantic Textons.

2) Visual Word Assignment,

where we compare a k-means visual vocabulary with a Random Forest and evaluate subsampling, dimension reduction with PCA, and division strategies of the Spatial Pyramid.

3) Classification,where we evaluate the RBF and Fast Histogram Intersection kernel for the SVM. Apart from the evaluation, we accelerate the calculation of densely sampled SIFT and SURF, accelerate nearest neighbor assignment, and improve accuracy of the Histogram Intersection kernel. We conclude by discussing whether further acceleration of the Bag-of-Words pipeline is possible. Our results lead to a 7-fold speed increase without accuracy loss, and a 70-fold speed increase with 3% accuracy loss. The latter system does classification in real-time, which opens up new applications for automatic concept classification. For example, this system permits five standard desktop PCs to automatically tag for 20 classes all images that are currently uploaded to Flickr.

1.6 SUSAN (Smallest Univalue Segment Assimilating Nucleus)

In order to retrieve a similarly look trademark from a large trademark database, an automatic content based trademark retrieval method using block hit statistic and corner Delaunay Triangulation features was proposed. The block features are derived from the hit statistic on a series of concentric ellipse. The corners are detected based on an enhanced SUSAN (Smallest Univalue Segment Assimilating Nucleus) algorithm and the Delaunay Triangulation of corner points are used as the corner features. Experiments have been conducted on the MPEG-7 Core Experiment CE-Shape-1 database of 1 400 images and a trademark database of 2000 images. The retrieval results are very encouraging.

2. Region Based Descriptor

In this section, three region-based shape descriptors, GMD, ZMD, and GD, are described and evaluated. GMD, ZMD, and GD are described and a comparison of the three shape descriptors is discussed and evaluation results are discussed.

2.1 Geometric Moment Descriptor (GMD)

The technique based on moment invariants for shape representation and similarity measure is extensively used in shape recognition. Moment invariants are derived from moments of shapes, and are invariant to 2D geometric transformations of shapes. The central moments of order $p + q$ of a two dimensional shape represented by function $f(x, y)$

2.2 Zernike Moment Descriptor (ZMD)

The use of orthogonal moments to recover the image from moments based on the theory of orthogonal polynomials, and introduced Zernike moments, which allow independent moment invariants to be constructed to an arbitrarily high order. The complex Zernike moments are derived from Zernike polynomials:

2.3 Grid Descriptor (GD)

The grid descriptor was proposed by Lu and Sajjanhar. It has been used in MARS and other applications. When Lu and Sajjanhar proposed the grid method, it was only applied to contour-based shape, and this convention is also followed by Chakrabarti et al. and Safar et al. In this section, it is improved to describe both contour and region shape.3.3.1 Grid Method In grid shape representation, a shape is projected onto a grid of fixed size, 16×16 grid cells, for example. The grid cells are assigned the value of 1 if they are covered by the shape (or covered beyond a threshold), and 0 if they are outside the shape. A shape number consisting of a binary sequence is created by scanning the grid in a left–right and top–bottom order, and this binary sequence is used as the shape descriptor to index the shape. For two shapes to be comparable using grid descriptors, several normalization processes have to be done to achieve scale, rotation, and translation invariance. A block diagram of computing grid descriptor for a contour-based shape is given in Fig.

It begins with finding out the major axis (MA), i.e., the line joining the two furthest points on the boundary. Rotation Angular radial transform (ART), which is the region-based shape descriptor of MPEG-7, has desirable properties for representing shape information in a small number of features with no redundancy. However, in order for ART to be useful, especially in limited computing environments, the computational cost of ART must be greatly reduced. In this paper, we derive symmetric/antisymmetric properties from the basis functions of ART and present a fast and efficient method to compute the ART coefficients using these properties. The proposed method significantly reduces the number of sinusoidal operations and multiplications in computing the coefficients of ART. Moreover, the memory requirements needed to store the ART basis functions

in lookup tables are only 25% of the conventional method. The experimental results are presented to show the effectiveness of the proposed method. Index Terms—Angular radial transform (ART), fast method,MPEG-7 shape descriptor, symmetry/antisymmetry.

The shape of objects in images provides a powerful visual clue for identification and recognition of the objects. Although other elementary visual features, such as color, texture, or motion, are equally important, they are usually not sufficient to reveal the object's identity. In particular, in binary images, the shape feature is the most important visual feature. Character recognition and trademark retrieval systems are typical examples of the use of such shape information .Angular radial transform (ART), which is the region-based shape descriptor of MPEG-7 , has desirable properties for representing shape information using a small number of features. Because it has orthogonal basis functions, ART can represent the shape of an object in an image with no redundancy or overlap of information between the coefficients. ART also has robustness to the noise and rotational invariance. ART employs a two-dimensional (2-D) complex transform de-finedona unit disk. Unfortunately, the transformation involves so many sinusoidal computations that, to date, no fast method has been reported to the best of our knowledge. The high computational complexity can be a constraint on some systems such as large image databases or limited computing environments. Therefore, minimization of the computational complexity of AR

2.4 Gabor

we present a new image enhancement algorithm based on Gabor wavelet coefficients fusion. The main idea of the algorithm: firstly we get a set of different directions Gabor features images, by filtering on the image with a set of Gabor wavelet filter; then integrate these Gabor features images under a certain scale by using the weighted average approach; and then, get edge image by thresholding; Finally, get the final enhanced image by adding the image using histogram equalization methods and edge image multiplied by a factor. Experiment results show that the method proposed by the paper can get a better enhanced image.

2.5 Curvlet Based

The research presented in this article is aimed at the development of an automated imaging system for classification of tissues in medical images obtained from computed tomography (CT) scans. The article focuses on using curvelet-based multi-resolution texture analysis. The approach consists of two steps: automatic extraction of the most discriminative texture features of regions of interest and creation of a classifier that automatically identifies the various tissues. The discriminating power of several curvelet-based texture descriptors are investigated. Tests indicate that energy, entropy, mean and standard deviation signatures are the most effective descriptors for curvelets, yielding accuracy rates in the 97-98% range. A comparison with a similar algorithm based on wavelet and ridgelet texture descriptors clearly shows that using curvelet-basedtexture features significantly improves the classification of normal tissues in CT scans

Texture feature plays a vital role in content based Image retrieval (CBIR). Wavelet texture feature modeled by generalized Gaussian density (GGD) [1] performs better than discrete wavelet texture feature. Curve let texturefeature was proposed in [2]. In this paper, we compute a new texture feature by applying the generalized Gaussian density to the distribution of curve let coefficients which we call curve let GGD texture feature. The purpose of this paper is to investigate curve let GGD texture feature and compare its retrieval performance with that of curve let, wavelet and wavelet GGD texture features. Experimental results show that both curve let and curve let GGD features perform significantly better than wavelet and wavelet GGD texture features. Among the two types of curve let based features, curve let feature shows better performance in CBIR than curve let GGD texture feature. The findings are discussed in the paper.

2.6 Dendrograms

A dendrogram that visualizes a clustering hierarchy is often integrated with a reorderable matrix for pattern identification.The method is widely used in many research fields including biology, geography, statistics, and data mining. However, most

dendrograms do not scale up well, particularly with respect to problems of graphical and cognitive information overload. This research proposes a strategy that links an overview dendrogram and a detail-view dendrogram, each integrated with a re-orderable

matrix. The overview displays only a user-controlled, limited number of nodes that represent the “skeleton” of a hierarchy. The detail view displays the sub-tree represented by a selected meta-node in the overview. The research presented here focuses on

constructing a concise overview dendrogram and its coordination with a detail view. The proposed method has the following benefits: dramatic alleviation of information overload, enhanced scalability and data abstraction quality on the dendrogram, and the

support of data exploration at arbitrary levels of detail. The contribution of the paper includes a new metric to measure the “importance” of nodes in a dendrogram; the method to construct the concise overview dendrogram from the dynamically-identified,

important nodes; and measure for evaluating the data abstraction quality for dendrograms. We evaluate and compare the proposed method to some related existing methods, and demonstrating how the proposed method can help users find interesting patterns through a case study on county-level U.S. cervical cancer mortality and demographic data.

2.7 Edge Direction Histogram

The H.264/AVC video coding standard aims to enable significantly improved compression performance compared to all existing video coding standards. In order to achieve this, a robustrate-distortion optimization (RDO) technique is employed to select the best coding mode and reference frame for each macroblock. As a result, the complexity and computation load increase drastically.This paper presents a fast mode decision algorithm for H.264/AVC intraprediction based on local edge information. Prior to intraprediction, an edge map is created and a local edge direction histogram is then established for each subblock. Based on the distribution of the edge direction histogram, only a small part of intraprediction modes are chosen for RDO calculation. Experimental results show that the fast intraprediction mode decision scheme increases the speed of intracoding significantly with negligible loss of peak signal-to-noise ratio.

4. Conclusion

Contour shape representations exploit only shape boundary information. Contour based methods gain popularity because it is usually simple to acquire and is descriptive sufficiently in many applications. There are generally two types of very different approach for contour shape modeling: conventional and structural. Conventional approaches treat the boundary as a whole, and a feature vector derived from the whole boundary is used to describe the shape. The measure of shape similarity is usually the Euclidean distance between the feature vectors. Structural approaches break the shape boundary into segments, known as primitives, using certain criterion. The final representation is usually a string or a tree, and the measure of shape similarity is string matching or graph matching.

In region based techniques all pixels with in a shape region are taken into account to obtain the shape representation. It uses moment descriptor to describe shape.

5. Bibliography

[1] C. Grigorescu, N. Petkov, and M.A. Westenberg, “Contour and Boundary Detection Improved by Surround Suppression of Texture Edges,” Image Vision and Computing, vol. 22, pp. 609-622, 2004.

[2] C. Grigorescu, N. Petkov, and M.Westenberg, “Contour Detection Based on Nonclassical Receptive Field Inhibition,” IEEE Trans.Image Processing, vol. 12, pp. 729-739, July 2003.

[3] A. Goshtaby, “Description and Discrimination of Planar Shapes Using Shape Matrix,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 7, pp. 738-743, Nov. 1985.

[4] L.J. Latecki and R. Laka ¨mper, “Shape Similarity Measure Based on Correspondence of Visual Parts,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, pp. 1185-1190, 2000.

[5] C. Grigorescu and N. Petkov, “Distance Sets for Shape Filters and Shape Recognition,” IEEE Trans. Image Processing, vol. 12, no. 10,pp. 1274-1286, 2003.

[6] G. Nagy, “Twenty Years of Document Image Analysis in PAMI,”IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 38-62, Jan. 2000.

[7] S. Belongie, J. Malik, and J. Puzicha, “Shape Matching and Object Recognition Using Shape Contexts,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 509-522, Apr. 2002.

[8] L.J. Latecki, R. Laka ¨mper, and U. Eckhardt, “Shape Descriptors for Non-Rigid Shapes with Single Closed Contour,” Proc. IEEE Conf.Computer Vision and Pattern Recognition, pp. 424-429, 1998.

[9] Y. Shelepin, O. Vahromeeva, A. Harauzov, S. Pronin, N. Foreman,and V. Chihman, “Recognition of Incomplete Contour and Half-Tone Figures,” Perception, vol. 33, supplement, p. 85c, 2004.

[10] L. 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,vol. 27, no. 1, pp. 142-147, Jan. 2005.

[11] A. Khotanzad and Y. H. Hong, “Invariant image recognition by Zernike moments,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 5, pp.489–497, 1990.

[12] S. X. Liao and M. Pawlak, “On the accuracy of zernike moments for image analysis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no.12, pp. 1358–1364, 1998.

[13] G. Friedland, K. Jantz, and R. Rojas, “Siox: Simple interactive object extraction in still images,” in IEEE Computer Soc. Int. Symp. Multi-media, Los Alamitos, CA, 2005, pp. 253–260.

[14] L. J. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in CVPR2000, Hilton Head Island, SC, Jun. 2000, vol. 1, pp. 424–429.