# Adaptive Feature Selection and Extraction Approaches for Image Retrieval based on Region

Haiyu Song
Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education,
Jilin University, Changchun, China;
Institute of Computer Graphics and Image Processing, Dalian Nationalities University, Dalian, China
Email: songhaiyu@gmail.com

Xiongfei Li*
Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education,
Jilin University, Changchun, China;
Email: lxf@jlu.edu.cn

Pengjie Wang
State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou, China
Institute of Computer Graphics and Image Processing, Dalian Nationalities University, Dalian, China
Email: wpj@dlnu.edu.cn

***Abstract*—Image retrieval based on region is one of the most promising and active research directions in recent year's CBIR, while region segmentation, feature selection and feature extraction of region are key issues. However, the existing approaches always adopt a uniform approach of segmentation and feature extraction for all images in the same system. In this paper, we propose adaptive image segmentation and feature extraction approach according to different category image for image retrieval system. To improve performance, we propose adaptive segmentation approach according to different category image. Textured image is segmented by Gaussian Mixture Models (GMM), while non-textured image is segmented by our proposed block-based normalized cut. To accurately describe feature of region, we propose weight assignment method for centroid pixel and its neighbor by convolution with normal distribution when image segmentation by GMM. To improve generalization, we propose adaptive number of Fourier descriptors of shape signature which depends on the energy distribution of Fourier descriptors, instead of fixed number by experience. To simply and efficiently describe the spatial relationships of multi-object or multi-region in same image, we apply simplified topological relationships. The experiments demonstrate that proposed approaches are superior to the traditional approaches.**

***Index Terms*—image retrieval, region-based, adaptive, feature selection and extraction, segmentation, Gaussian Mixture Models, Fourier descriptor, normalized cut**

## I. INTRODUCTION

The advances of computing and multimedia technologies have led to an accumulation immense multimedia data, especially image data. Consequently, how to retrieve similar image is becoming a challenge. In order to solve this problem, researchers first propose text-based image retrieval, which retrieves image according to query word. The approach is time consuming and subjective, moreover, it can't process images without any associated texts. In order to overcome these drawbacks, recent researchers on image retrieval focus on content-based image retrieval (CBIR), which describes image content with low level image features such as color, texture, and shape. Here, "content" is some kind of objective statistic character of images. Most CBIR systems utilize global visual statistic information, which couldn't be understood directly by human beings. Usually, there is a deep semantic gap between low-level visual content and high-level semantic concept. It is hard for global features to reduce the semantic gap. Local features often correspond with more meaningful image components such as objects and entities, which make association of semantics with image portions straightforward. As the result, we have witnessed a shift from global feature representations for images such as global color histogram and global shape descriptors to local features and descriptors such as salient points and SIFT in resent years.

Global descriptors can't imply semantic object, while local descriptors is sensitive to noise. Region-based descriptor is a compromise between global and local descriptors, consequently, more and more researchers focus on region-based image retrieval, for which segmentation is its first step. After segmentation, features are extracted from segment or regions instead of whole image, that is to say, features are extracted from each region of original image. However, the existing region-based image retrieval approaches segment image and extract feature usually in a uniform manner for all the images in the same system. In fact, it is hard to find a

Manuscript received January 1, 2009; revised February 20, 2009; accepted July 1, 2009
* Xiongfei Li is corresponding author.


doi:10.4304/jmm.5.1.85-92

single segmentation or feature extraction approach suitable to all images. Motivated by the efficiency and performance of region-base image retrieval, we propose an adaptive feature selection and extraction approach, which is completely data-driven without any prior knowledge. The proposed approach adaptively applies different segmentation method to different kind of image after image classification by SVM, and the number of shape Fourier descriptors of each image region depends on its energy distribution instead of arbitrary or prior fixed number.

The remainder of this paper is organized as follows: Section 2 introduces the system design; Section 3 presents adaptive image segmentation; Section 4 presents feature extraction of color and texture; Section 5 presents shape feature extraction; Section 6 presents spatial relationship extraction; Section 7 presents dimension reduction method; Section 8 presents experiments and results; Section 9 gives conclusions and future work.

## II. System Overview

This section gives an overview of our system as shown in figure 1. When a new image is stored into database, or a user submits a query request by sample image, the system decides if the image is textured image by a classifier implemented by SVM. Then, the system applies suitable segmentation approach for each category. We use improved N-cut algorithm to segment non-textured image, while Gaussian Mixture Models (GMM) to textured image. Construct normalized histogram as color feature vector of region, $F_C$, n-order moment of histogram as texture vector of region, $F_T$. We use centroid distance as signature function of shape, and compute its Fourier descriptors by Fourier transform. With the aim of improving generalization, we propose an adaptive algorithm to determine the number of Fourier descriptors of shape signature which depends on the energy distribution of FDs instead of fixed number by experience. To simply and efficiently describe the spatial relationships of multi-object or multi-region in same image, we apply simplified topological relationships, and extract spatial feature vector by computing graph spectra of spatial relationship. The comprehensive feature vector of region k, $Fe_K$, is composed of color feature vector $F_C$, texture feature vector $F_T$, shape feature vector $F_S$, and spatial feature vector $F_{Sp}$. That is to say, $Fe_K=\{F_C, F_T, F_S, F_{Sp}\}$. To improve the retrieval efficiency and information discrimination, we apply Isomap to reduce dimension of feature vector. For region-based feature, feature vector corresponding to each image has different size. We use Earth Movers Distance as similarity measurement instead of traditional Euclidean distance.

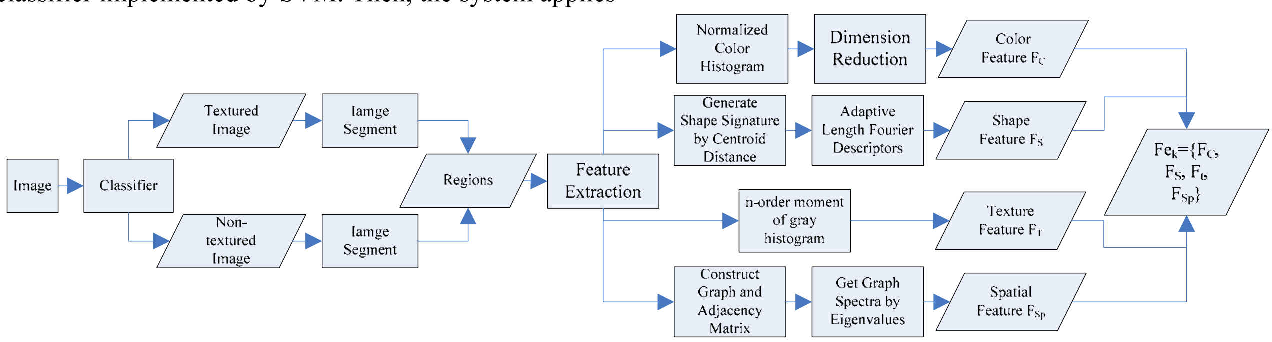



Figure 1. The overview of proposed system /approaches.

## III. . Image Segmentation

Image segmentation is a key step to acquire a region-based signature. Shape signature or shape similarity is meaningless without reliable segmentation. There are many kinds of segmentation approaches, such as Canny operator, k-means clustering. All operators such as Canny operator utilize local information to filter, which couldn't ensure a continuous closure border. To construct closure border shape signature, the most widely used segmentation approach is k-means clustering, whose advantage is high speed.

There is not any segmentation approach suitable for all categories image. Experiments show that clustering algorithms are suitable for non-textured image, while GMM is suitable for textured image. To improve speed, we propose block-based normalized cut based on normalized cut image segmentation algorithm. To improve performance, we propose weighted centroid pixel gray value as one of the region feature values by convolution with normal distribution when use GMM as segmentation approach for textured image.

### *A. Normalized Cut*

Cut segmentation is a new advance in this field, which is an application of spectral clustering to image segmentation. The algorithm treats image segmentation as a graph partition problem. The Cut algorithm maps image segmentation problem to graph partition. Any image can be represented as a weighted undirected graph G= (V, E), where the nodes of the graph are the points in the feature space, and an edge is formed between every pair of nodes. A graph G=(V, E) can be partitioned into two disjoint sets by simply removing edges connecting the two parts. The degree of dissimilarity between these two pieces can be computed as total weight of the edges that have been removed. In graph theoretic language, it is called the cut:

$$\text{cut}(A, B)=\sum_{u\in A, v\in B} w(u,v) \ .$$

The optimal partitioning of a graph is the one that minimizes the cut value. The minimum cut criteria favors cutting small sets of isolated nodes in the graph [1]. To avoid unnatural bias for partitioning out small sets of points, Jianbo Shi proposed a new measure of disassociation between two groups [2]. Instead of looking at the value of total edge weight connecting the two partitions, the measure computes the cut cost as a fraction of the total edge connections to all the nodes in the graph, and the disassociation measure was called the Normalized Cut (N-cut):

$$\text{N-cut (A, B)} = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)},$$

where assoc(A, V)= $\sum_{u\in A,t\in V} w(u,t)$ is the total connection from nodes in A to all nodes in the graph and assoc(B, V)is similarly defined. Compared with previous clustering segmentation algorithms, the N-cut aims at extracting the global impression of an image, and measuring both the total dissimilarity between the different groups as well as the total similarity within the groups. Because N-cut comprehensively considers global and local information as well as its robustness, it can achieve better performance than previous clustering algorithms for image segmentation.

### *B. Improved Normalized Cut*

Every vertex of graph is corresponding to a pixel of image. The N-Cut looks like considerably appealing because it considers region segmentation from global perspective. It can be used to pattern recognition and other domain, but it is infeasible for online image retrieval due to memory and time complexity as considering medium and large datasets. When an image is mapped to a graph, the graph always is described by matrix, whose element number is (MxN)x(MxN) if the image size is MxN. As the result, the incredible computational cost makes it lost usefulness.

We propose an improved N-Cut algorithm, which is not based on single pixel but 8x8 pixel image block when image problem converted to graph. In the proposed N-Cut algorithm, each node of graph is corresponded to a image block instead of a single pixel. The weight on each edge, $w_{ij}$ between node i and j of adjacency matrix W, is a product of a feature similarity term and spatial proximity term: $w_{ij}=\exp\{-\frac{\|F(i)-F(j)\|_2^2}{\delta_I}\}*\exp\{-\frac{\|X(i)-X(j)\|_2^2}{\delta_X}\}$, if $\|X(i)-X(j)\|_2 < r$; otherwise, $w_{ij}=0$, where X(i) is the spatial location of node i, and F(i) is a feature vector based color information at the node i. When the formula is used in the standard N-cut algorithm, the F(i) always represents the intensity value of corresponding pixel. While when it applied in our block-based N-cut, the F(i) represents a visual feature vector of corresponding image block. Because the N-cut algorithm is clustering of adjacent pixels, the weight of node i and j is 0 when the distance of two nodes is larger than the threshold r. We select the color feature as feature vector of image block, due to the block is homogeneous through N-cut segmentation.

Our proposed block-based N-cut algorithm is as follows:

a) Segment image into several 8x8 pixel blocks
b) Construct graph G, whose vertex is corresponding to image block.
c) Compute weighted adjacency matrix W, whose edge weight $w_{ij}$ is similarity and distance metric of node i and j corresponding to block i and j respectively.
d) Compute the unnormalized Laplacian L.
e) Compute the first k eigenvectors $v_1, \ldots, v_k$ of the generalized eigenproblem $Lv = \lambda D_v$.
f) Let $V \in R^{n\times k}$ be the matrix containing the vectors $v_1, \ldots, v_k$ as columns.
g) For $i = 1, \ldots, n$, let $y_i \in R^k$ be the vector corresponding to the i-th row of V .
h) Cluster the points $(yi)_{i=1,\ldots,n}$ in $R^k$ with the k-means algorithm into clusters $C_1, \ldots, C_k$.

If the size of image is MxN, the size of adjacent matrix of original N-Cut is MxNxMxN while improved N-Cut is (M/8)x(N/8)x(M/8)x(N/8). Compared with the original N-cut algorithm, improved algorithm can greatly decrease computational cost (1/4096) with a little loss of accuracy.

### *C. Gaussian Mixture Model Segmentation*

After segmented using clustering method, the pixels of image region are similar in color, while pixels of the same textured region always are dissimilar in gray values or density, so the clustering algorithms are not suitable for texture images. For textured image, we adopt Gaussian Mixture Model (GMM) to segment image. In this algorithm, we use 5 statistics as feature vector of image block when trained and identified in GMM, including average, variance, maximum, minimum, and weighted centroid pixel gray value. We think the centroid pixel value has a important role in deciding the block category and the neighbor pixels has a certain effect on it, so weighted centroid pixel value is calculated through convolution operator by normal distribution. Suppose the distribution function of pixels intensity in block is f, and the normal distribution is g, the weighted centroid pixel value is the convolution product of f and g, namely f*g.

The steps of algorithm are as follows.

a) Preprocess the raw image.
b) Randomly select N 8x8 pixel blocks as training sample.
c) Construct feature vector of 8x8 pixel block, including average, variance, maximum, minimum, and weighted centroid pixel. Weighted centroid pixel is calculated by convolution product of pixels value distribution and normal distribution.
d) Initialize parameter values, such as the number of category, expectation, and variance of Gaussian Mixture Models.
e) Estimate parameter using EM algorithm, and iterate the process until convergence.
f) Compute posterior probability. Compute

posterior probability that each image block falls into corresponding class according to likelihood estimation.

g) Image segmentation. Determine the texture category of each image block according to posterior probability so as to perform image segmentation.

## IV. Feature Extraction of Color and Texture

### A. Color Feature Extraction

Color is basic information of image content, and the most widely used visual feature for image retrieval due to the computational efficiency of feature extraction, robust to the noise, invariant to resolution, orientation, resizing, and translation. A typical description for color information is the color histogram, which describes the color distribution of an image in a simple and efficient method. All colors can be represented by variable combinations of the three components: red(R), green (G), and blue (B), that is can be represented in RGB color space. There are some other color spaces, such as HSV, L*u*v*, YIQ, etc. RBG is suitable for device, while HSV is suitable for human perception. We use HSV as color space. Transforming the color image from RGB space to HSV space, and then quantizing the triple-color components (H, S, V) into non-equal intervals. The transformation formula is described as follows [3]:

$$H=\begin{cases}0 & \text{if } h\in[316,20]\\ 1 & \text{if } h\in[21,40]\\ 2 & \text{if } h\in[41,75]\\ 3 & \text{if } h\in[76,155]\\ 4 & \text{if } h\in[156,190]\\ 5 & \text{if } h\in[191,270]\\ 6 & \text{if } h\in[271,295]\\ 7 & \text{if } h\in[296,315]\end{cases},\quad S=\begin{cases}0 & \text{if } s\in[0,0.2]\\ 1 & \text{if } s\in(0.2,0.7]\\ 2 & \text{if } s\in(0.7,1]\end{cases},\quad V=\begin{cases}0 & \text{if } v\in[0,0.2]\\ 1 & \text{if } v\in(0.2,0.7]\\ 2 & \text{if } v\in(0.7,1]\end{cases}$$

To reduce the computation complexity and storage space, we map the triple-color components to a one-dimension vector by formula L=9H+3S+V. Therefore the color feature of image can be represented by a vector, whose element value range is [0, 71]. Consequently, each region of an image can be represented a weighted feature vector by a normalized 72-bin color histogram. If the feature vector only includes color feature, the weighted feature vector for region k can be represented as $F_c=[h_0,\ldots h_{71}, f_k]$, $f_k$ representing the percentage of pixels falling into the region k.

### B. Texture Feature Extraction

Texture is a very important visual cue for identifying regions with homogeneous periodic patterns, and it has been shown that texture features have a very high discriminating power. An important and first step to identify the perceived texture feature is to build mathematical models representing the intensity variations in an image. There are many methods to describe texture feature, such as co-occurrence matrix, texture structure, frequency spectral method, and gray histogram. We compare the above methods by experiments. The results show that frequency spectral method, especially wavelet, gets the best effect, but it is not suitable for online image retrieval due to its expensive computational cost. Co-occurrence matrix method gets better effect, but it is high computational complexity to construct co-occurrence matrix. To meet large-scale data-processing needs, we adopt gray histogram, whose algorithm is simple and easy to implement. Its texture feature vector is constructed based on histogram by computing moments, whose steps are as follows:

a) Compute the average of gray histogram, $m=\sum_{i=0}^{71} x_i f(x_i)$

b) Compute n-order moment of histogram (n=1, 2, 3, 4) , $u_n(x)=\sum_{i=1}^{71}(x_i-m)^n f(x_i)$

c) Construct feature vector, $F_T=[u_1, u_2, u_3, u_4]$

As the region is homogeneous of color and texture, the feature vector $F_T$ can be computed from region k directly. When we extract feature $F_T$ representing texture feature of region k, to reduce computational complexity, we can compute $F_T$ only from inner pixels of rectangle inscribed in region k instead of from whole region k.

## V. Shape Feature Extraction

Shape is one the most important low level image feature due to that shape is a very important feature to human perception. Among the color, texture, and shape, the shape is the best attribute that associates with semantic object. Shape representation methods can be classified into two categories: region based and contour based. Compared with region-based method, contour based shape representation is more popular and suitable for image retrieval.

We compared those shape signatures suitable for image retrieval such as centroid distance, complex coordinates, curvatures function, and cumulative angles. The experiments demonstrate that centroid distance is best signature function among the four, while the cumulative angles algorithm is worst. The reason can be explained that the centroid distance not only captures boundary information but also region information, that is to say, it captures both local and global shape features. In a conclusion, centroid distance is a desirable shape signature for image retrieval based on region.

Shape signatures require intensive computational cost during similarity calculation, due to the hard normalization of rotation invariance. As the result, these shape signatures can be conveniently used in image retrieval only after processed by Fourier descriptors (FD). The Fourier transformed coefficients form the Fourier descriptors of the shape. These descriptors represent the shape of the object in a frequency domain. The lower frequency descriptors contain information about the general features of the shape, and the higher frequency descriptors contain information about finer details of the shape. The very high frequency information describes the small details of the shape, it is not so helpful in shape discrimination, therefore, and they can be ignored. As the result, the dimensions of the Fourier descriptors used for indexing shapes are significantly reduced and the feature vector is easy to be normalized.

For the discrete Fourier transform, Parseval's theorem can be often written as: $\sum_{i=0}^{N-1}|r[i]|^2 = \frac{1}{N}\sum_{i=0}^{N-1}|FD_i|^2$. As centroid distance is a real-valued function, hence there are only N/2 different DFT coefficients, we can get the first-half as Fourier Descriptors, such as FDs= $\left[|FD_0|,|FD_1|,|FD_2|,\ldots,|FD_{N/2}|\right]$, which can be used to measure similarity without any error. Consequently, simplified energy formula is written as: $E_{simplified}=\frac{1}{N}\sum_{i=0}^{N/2}|FD_i|^2$.

To compare completely different signature and FDs, the traditional method truncating the number of FDs always bases on the arbitrary decision or the prior experiences. The number of FDs according to the way can hardly reach a better approximation to the descriptions of all the images. To improve generalization, we propose an adaptive algorithm to determine the number of Fourier descriptors of shape signature which depends on the energy distribution of FDs, instead of fixed number by experience. We can set a threshold as approximation to all energy of FDs in advance, such as $T_{Energy}$. Suppose $E_{simplified}$ (T)= $\sum_{i=0}^{T-1}|FD_i|^2$, we can get truncated T by solving inequality $\frac{\sum_{i=0}^{T}|FD_i|^2}{\sum_{i=0}^{N/2}|FD_i|^2} > T_{Energy}$.

Then we can get truncated FDs= $\left[|FD_0|,|FD_1|,|FD_2|,\ldots,|FD_T|\right]$.

The steps of extracting shape feature vector are as follows.

a) Calculate the centroid ($x_c$, $y_c$) of the shape, $x_c=\frac{1}{L}\sum_{t=0}^{L-1}x(t)$, $y_c=\frac{1}{L}\sum_{t=0}^{L-1}y(t)$.

b) Get shape signature by centroid distance, $r(t)=([x(t)-x_c]^2+[y(t)-y_c]^2)^{1/2}$

c) Perform discrete Fourier transform on centroid distance shape signature to get Fourier Descriptors (FDs), which presented as $FD_n$, n=0,1,…N-1.

d) As centroid distance is real value, there are only N/2 different DFT coefficients, we can get the first-half as Fourier Descriptors, then we can get FDs: FDs= $\left[|FD_0|,|FD_1|,|FD_2|,\ldots,|FD_{N/2}|\right]$, which can achieve rotation invariant of FDs by ignoring the phase information.

e) Get truncated T by solving inequality $\frac{\sum_{i=0}^{T}|FD_i|^2}{\sum_{i=0}^{N/2}|FD_i|^2} > T_{Energy}$

f) Construct shape feature vector, $F_{shape}=\left[\frac{|FD_1|}{|FD_0|},\frac{|FD_2|}{|FD_0|},\ldots,\frac{|FD_T|}{|FD_0|}\right]$

## VI. SPATIAL RELATIONSHIP EXTRACTION

Spatial relationship is important visual feature for multi-object image retrieval. Spatial relationship mainly includes direction relationship and topological relationship. The latter is suitable for CBIR.

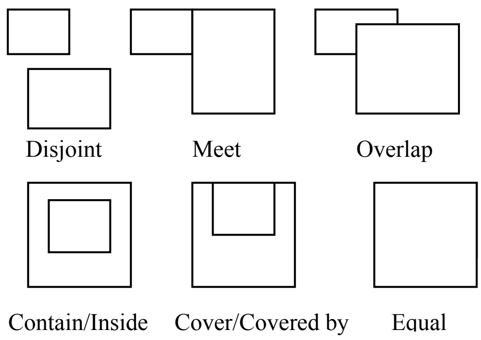



Figure 2. Traditional topological relationships.

Egenhofer uses disjoint, contain, inside, cover, covered by, equal, overlap, and meet these eight kinds of relationships to describe object spatial topological relationship. Topological relationships are shown in Figure 2.

### *A. Simplified Topological Relationships*

The topological relationship model proposed by Egenhofer is a more precise, comprehensive and expected way to describe topological relationship, and is widely accepted. But the over-detailed division of topological relationship is not practical for image retrieval. In reference [4], we propose the simplified topological relationships. To efficiently describe the topological relationship of objects, most systems use the minimum bounding rectangle (MBR) to represent the object approximately, which always lead to poor precision and

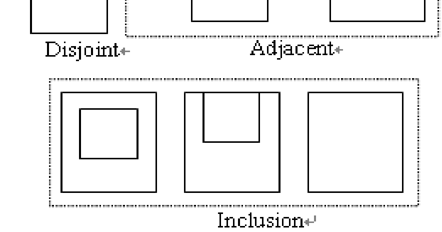



Figure 3. Simplified topological relationships.

low recall of retrieval result. We proposed simplified topological relationship. As it shown in figure 3, the traditional topological relationships can be simplified as three categories: disjoint, adjacent, and inclusion. Meet and overlap are contained in adjacent, while other 5 relationships, such as contain, inside, cover, covered by and equal, are contained in inclusion.

### *B. Topology Feature Vector Extraction*

We use graph spectra to represent topological relationships. The graph spectra calculating process is as follows [4].

a) Generate topology graph according to object topological relationships in image.

b) Generate adjacency matrix by topology graph.

c) Get the eigenvalues of adjacency matrix.

d) Get their absolute values and sort them in descending order, to get the corresponding graph spectra.

e) Construct feature vector of topological relationship by the n largest elements of graph spectra.

The region after segmented by clustering algorithm is relatively homogeneous. Suppose the number of clusters is n, the image will be segmented into n clusters. Let the mean of feature vectors in the cluster k be $Fe_k$, and the pixel number of cluster k is $f_k$ percent of number of whole image. The signature of image is $\{(Fe_1, f_1), \ldots, (Fe_k, f_k), \ldots, (Fe_n, f_n)\}$, which is called region-based signature, where n depends on processing image. The region-based feature of each image can be regarded as sets of weighted vectors, where $f_k$ means the weight of region k. The comprehensive feature vector of region k, $Fe_K=\{F_C, F_T, F_S, F_{Sp}\}$, where $F_C$ represents normalized color feature vector, $F_T$ represents texture feature vector, $F_S$ represents shape feature vector, and $F_{Sp}$ represents spatial relationship relationships of region k.

## VII. Dimension Reduction

To avoid the curse of the dimensionality, we apply dimension reduction technology into color feature vector for its high dimension in our feature selection and extraction.

Compared with traditional linear technology, manifold learning can discovery the intrinsic dimensions of nonlinear high-dimensional data effectively, helping researchers to reduce dimensionality and analyze data better. We choose Isomap as dimensionality reduction of color feature vector. Isomap is a nonlinear dimensionality reduction technique that uses MDS techniques with geodesic interpoint distances instead of Euclidean distances [5]. Geodesic distances represent the shortest paths along the curved surface of the manifold. Unlike the linear techniques, Isomap can discovery the nonlinear degrees of freedom that underlie complex natural observations [5]. We can map 72-dimension color feature to 8-dimension by Isomap, whose algorithm is as follows:

a) Determine which points are neighbors on the manifold M, based on the distances between pairs of points in the input space.

b) Create a graph G, with edges between neighbors and distance weights

c) Estimate the geodesic distances between all pairs of points on the manifold M by computing all-pairs shortest paths in graph G.

d) Apply classical MDS to matrix of graph distances, constructing an embedding of the data in a k-dimensional Euclidean space Y that best preserves the manifold's estimated intrinsic geometry.

## VIII. Experiments and Results

We have implemented these improved approaches in a prototype system. To test improved approaches, we download 5000 natural images from Internet as evaluation dataset. The experiments are performed on the dataset, which include 50 categories and each category includes 100 images.

We apply adaptive segmentation approach to any image firstly, then, construct color feature vector, texture feature vector, shape feature vector, and spatial relationship feature vector. We compare our proposed approaches, namely block-based N-cut / GMM, the weight centroid pixel value, adaptive number of FDs for shape signature, simplified spatial relationships, with the state-of-the-art approaches. The experiments show our proposed approaches are superior to the traditional approaches.

To evaluate the performance of image segmentation, we first applied GMM and N-cut to all images, then we applied adaptive method, that is to say, GMM for textured image and improved N-cut. The speed of our improved N-cut is 6-9 times than that of the classic method with sacrificing a little segmentation quality.

### *A. Similarity Computation*

Widely used similarity computation methods, such as Euclidean distance, Minkowski, K-L distance, are one to one mapping, which can only be used to measure similarity between two vectors with same size. For feature vector based on region, its size corresponding to every region is same. But mages have different number of regions. So we can't compute similarity between two images by one to one mapping. Considering the weight of region-based feature vector, we use Earth Movers Distance (EMD) as similarity measurement [6]. The definition of the distance is thus

$$\mathrm{EMD}(I_1, I_2)=\frac{\min\limits_{s_{i,j}} \sum_{i=1}^{n1}\sum_{j=1}^{n2} s_{i,j} d(z_i^{(1)}, z_j^{(2)})}{\sum_{i=1}^{n1}\sum_{j=1}^{n2} s_{i,j}},$$

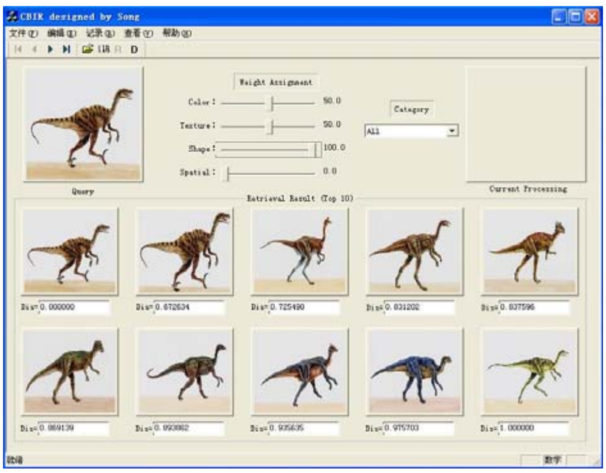
where $n_1$, $n_2$ are the number of region segmented from image $I_1$ and $I_2$ respectively, while $z_i^{(1)}$ are the feature vector representing region i of image $I_1$, and as well as $z_j^{(2)}$. The weight $s_{i,j}$ indicates the significance of associating $z_i^{(1)}$ with $z_j^{(2)}$.

Our adaptive method determining the number of FDs of shape signature achieved a better generalization due to it depending on energy distribution of FDs instead of fixed number by experience. In fact, the adaptive method is independent of visual features, so it can be used to FDs of color or texture feature.

### *B. Evaluation Metrics*

The two most popular evaluation measures are: precision and recall. Precision is the percentage of retrieved images that are relevant to the query image. Recall is the percentage of all the relevant images in the search database which are retrieved. The shortcoming of precision is that it is calculated for the entire retrieved set and is unaffected by the respective rankings of the relevant entities in the retrieved lists [7]. Mean average precision (MAP) is the arithmetic mean of average precision calculated over a number of different queries, while average precision is given as the mean of all the individual precision scores [8]. We use MAP and recall as evaluation metrics.

(A). Fixed number of FDs.



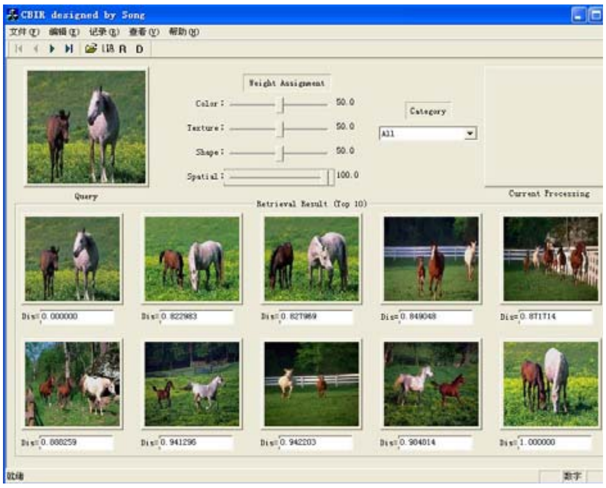
(B). Adaptive number of FDs.

Figure 4. Region-based retrieval with adaptive number of FDs compared with fixed number, shape feature with bigger weight

### *C. Results*

We perform experiments to compare our adaptive determining the number of FDs with traditional fixed number method. As shown in figure 4, the precision in adaptive number (B) is better than fixed number (A).

(A). Global feature including color and texture feature



(B). Region-based feature including color, texture, shape, spatial relationship

Figure 5. Retrieval performance comparing global feature with region-based feature

As shown in figure 5, our proposed approach can get spatial relationships from query image, and retrieve similar images not only in color, texture of region, shape, but also in spatial relationships of regions. Our proposed approach can imply semantic concept or object, and complete sub-image query, which is extremely meaningful to narrow the semantic gaps between low-level content and high-level concept.

Eexperiment results show that our proposed approaches result in 32% average improvement over traditional approaches according to MAP evaluation metric. Overall, our improved N-cut and GMM, adaptive number of FDs, and simplified topological relationships are the key performance contributors for the system.

## IX. CONCLUSIONS AND FUTURE WORK

In this paper, we propose adaptive image segmentation for textured or non-textured image. To increase speed so as to make N-cut as a practical and valuable method, we propose block-based N-cut spectral clustering as image segmentation method. To improve generalization, we propose adaptively determining the number of FDs of shape signature depends on its energy distribution. Our simplified topological relationships approach is a simple model, but it has high discriminative power. The results have shown that our adaptive feature selection and extraction approaches are good choice for image retrieval, especially for region-based image retrieval.

To further improve the segmentation quality, we can use k-means as preprocessing of block-based normalized cut algorithm instead of fixed size block, such as 8x8 pixel block. Even if the selection and extraction approaches of feature are optimal, there are semantic gaps between low-level visual content and high-level concept. Automatic image annotation is the most promising technology to narrow the semantic gap. We believe that automatic image annotation will be research focus in image retrieval, analysis and understanding domain in the future. Moreover, image retrieval will merge more other domain knowledge, such as machine learning, pattern recognition, computer vision, relevance feedback, and human-computer interaction. The overall goal is to bridge the semantic and sensorial gaps using the available visual features of images and relevant domain knowledge [8].

## ACKNOWLEDGMENT

We thank Lintong Huang and Xiaoli Zhang for their helping in experiment and evaluation. This work was supported in part by National Natural Science Foundation of China under grants 60675008, and Youth Foundation of Dalian Nationalities University under grants 2006A203.

## REFERENCES

[1] Z. Wu, R. Leahy. "An optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol.15, pp.1101-1113, Nov. 1993

[2] Jianbo Shi, Jitendra Malik. "Normalized Cuts and Image Segmentation", *IEEE Transaction on Patterm Analysis and Machine Intelligence*. vol.22, pp.888-95, August 2000.

[3] WAN Hua-lin, Morshed U. Chowdhury. "Image Semantic Classification by Using SVM", *Chinese Journal of Software*, vol.14, pp.1891-1899, November 2003.

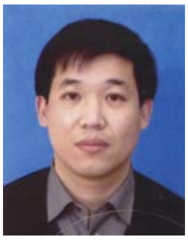
[4] Song Haiyu, Wang Pengjie, Li Xiongfei. "Research and Design of Drawings Retrieval System based on Spatial Relationships and Shape", *Proceeding of 2008 ISECS International Colloquium on CCCM*, Los Alamitos: IEEE Computer Society Press, pp. 336-341, August 2008.

[5] J. Tenenbaum, V. de Silva, J. Langford. "A Global Geometric Framework for Nonlinear Dimensionality Reduction", *Science*, vol.290, pp.2319-2323, Dec.2000

[6] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. "The earth mover's distance as a metric for image retrieval", *Internal Journal Computer Vision*, vol.40, pp.99-121, 2000.

[7] Huijsmans, D. P. and SEBE, N. "How to complete performance graphs in content-based image retrieval: Add generality and normalize scope", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, pp. 245-251, February 2005.

[8] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. "Image Retrieval: Ideas, Influences, and Trends of the New Age", *ACM Transactions on Computing Surveys*, vol.40, , pp.5:1-60, April 2008



**Haiyu Song** received the BS degree in computer & application in 1996, the MSc degree in computer software and theory in 2003, all from Jilin University, china. Now he is a Ph.D. Candidate in computer software and theory at Jilin University.

During 1996-2000, he was a software engineer in ACBC Co., Ltd., china, where the team chaired by him has developed a data acquisition and management system for PHILIPS gauge, rejected product automatic recognition system, and other pattern recognition software. Since 2003, he has been a member of the faculty of the computer science and engineering at Dalian Nationalities University, Dalian, China. He has authored or coauthored more than 20 research papers, such as "A General Multi-step Algorithm for Voxel Traversing Along a Line" (*Computer Graphics Form*, vol27(1), 2008), "Study on Model and Method of Image Data Mining" (*Journal of Jilin University*, vol32(1), 2002). His research interests include image analysis and understanding, image retrieval, data mining, and computer graphics.

**Xiongfei Li** received the BS degree in computer software in 1985 from Nanjin University, the MSc degree in computer software in 1988 from the Chinese academy of sciences, the PhD degree in communication and information system in 2002 from Jilin University.

Since 1988, he has been a member of the faculty of the computer science and technology at Jilin University, Changchun, China. He is a professor of computer software and theory at Jilin University. He has authored more than 40 research papers, including "A Data Mining Algorithm Based on Calculating Multi-Segment Support"(*Chinese Journal of Computer*, vol24(6), 2001), "An Algorithm for Discovering Association Rules Based on the Attributes of Items"(*Chinese Journal of Computer*, vol25(12), 2002). His research interests include data mining, intelligent network, and multimedia.

Prof. Li is a member of the IEEE. He is also an expert of minister of education evaluating invite tender for electromechanical product.

**Pengjie Wang** received the BS degree in computer science and technology in 2001, the MSc degree in computer software and theory in 2004, all from Jilin University, china. Now he is a Ph.D. Candidate in computer software and theory at Zhejiang University.

Since 2003, he has been a member of the faculty of the computer science and engineering at Dalian Nationalities University, Dalian, China. He has authored or coauthored more than 20 research papers, including "Improved point-by-point algorithm for generating nonparametric curves" (*Journal of Computational Information Systems*, vol3 (2), 2007). His research interests include computer graphics, and digital entertainment.