Volume : 4, Issue: 1
January - June 2014

ISSN : 2229 - 3515

International Journal of ADVANCES IN SOFT COMPUTING TECHNOLOGY

Editor-in-Chief
Dr. Vaka Murali Mohan

Published by
BHAVANA RESEARCH CENTER

Feature Identification Using Combined ICA - Wavelet Method for Image Mining

C. Srinivas[1] and Dr. S. K. Mohan Rao[2]

1. Assoc. Prof., Dept. of CSE, Sree Visvesvaraya Institute of Tech. & Science, Mahabubnagar, AP
2 Principal & Professor, Dept. of CSE, Siddhartha Institute of Science & Tech., Hyderabad, AP

KEYWORDS

Independent component analysis (ICA);

Feature Fusion;

Image Information Mining;

Wavelets;

Abstract: *In this paper we proposed here shows that a wavelet-based technique reduces the complexity and dimensionality of the extracted features. This reduced set of features can help to expedite the image retrieval in Earth observation data archives. In general, image transformation is a primary part in the image information mining and knowledge discovery process, which assists in the discovery of new and interesting relationships between the features. Instead of using the raw image, transformation of datasets is used to improve the quality of knowledge discovery in an image. The transform of the data can take place in many different forms, i.e., applying arithmetic operation (+,-,*) on a feature set, combining nonlinearly correlated features, the transformation along the spectral axis to obtain new features (RGB to HSV space), linear transformation - Principal component analysis (PCA), non-linear transformation –Independent component analysis (ICA), and denoising features using wavelet transforms. ICA, a variant of principal component analysis (PCA), provides a linear transformation that minimizes statistical dependence. ICA removes the rotational invariance of PCA and also provides components that are uncorrelated and statistically independent. ICA has been successfully used to solve object recognition and classification problems. ICA is also gaining popularity in the remote sensing field. In particular, ICA has been successfully applied for unsupervised classification of hyper spectral imagery. The image segmentation process can be performed speedily by reducing the amount of raw pixel data used for unsupervised classification. The wavelet transform decomposes an image into multiple sub-bands with different frequencies. The coefficients at a coarse level (LL sub-band) can be used for coarse-scale image segmentation, which results in reduction of raw pixel data used for unsupervised classification. Region identification coefficients from different frequency sub-bands are used to capture the texture detail of the region. Results show that the presented feature set is computationally less expensive and more efficient in capturing the spectral and spatial texture information. After extensive experimentation with different types of mother wavelets, it can be concluded that reverse Biorthogonal wavelets of shorter length and the simple Haar filter provided better results for the image information mining from the database used in this study.*

1. INTRODUCTION

The Computer Industry has seen a large growth in technology – access, storage and processing fields. This combined with the fact that there are a lot of data to be processed has paved the way for analyzing and mining data to derive potentially useful information. Various fields ranging from Commercial to Military want to analyze data in an efficient and fast manner. Particularly in the area of Multimedia data, images have the stronghold. However there is a general agreement that sufficient tools are not available for analysis of images. One of the issues is the effective identification of features in the images and the other one is extracting them. One of the difficult tasks is that to knowing the image domain and obtaining a priori knowledge of what information is required from the image.

*** Mr. C. SRINIVAS**
Associate Professor
Dept. of Computer Science & Engg.
Sree Visvesvaraya Institute of Tech. & Science
Mahabubnagar, AP, India
Ph. No: 91-9849715531
E- Mail: sharanasri1@gmail.com

This is one of the reasons the image mining process cannot be completely automated.

Image mining deals with extraction of implicit knowledge, image data relationship or other patterns not explicitly stored in images and uses ideas from computer vision, image processing, image retrieval, data mining, machine learning, databases and AI.

The police maintain image database of criminals, crime scenes and stolen items. In the medical profession X-rays, and scanned image database are kept for diagnosis, monitoring and research purposes. In architectural and engineering design image database exist for design projects, finished projects and machine parts. Features based on color, texture (Harlick features), and shape are extracted for each region and indexed in the database [1]. These features fail to capture the local properties of regions and are also computationally expensive in terms of query retrieval. Recent research has shown that a wavelet-based technique reduces the complexity and dimensionality of the extracted features [2-3]. The transform of the data can take place in many different forms, i.e., applying arithmetic operation (+,-,*) on a feature set, combining nonlinearly correlated features, the transformation along the spectral axis to obtain new features (RGB to HSV space),linear transformation - Principal component analysis (PCA), non-linear transformation –Independent component analysis (ICA), and de-noising features using wavelet transforms [4].

2. MODEL

In small collection of images simple browsing can identify an image. This is not the case for large and varied collection of images, where the user encounters image retrieval problem. A typical retrieval problem example is a design engineer who needs to search his organization database for design projects similar to that required by his clients or the police seeking to confirm the face of a suspected criminal among faces in the database of renowned criminals.

2.1 Visual content levels

Images are naturally endowed with attributes or information content that can help in resolving the image retrieval problem. The information content that can be derived from an image is classified into three levels. See Figure1.

• Low level: They include visual features such as colour, texture, shape, spatial information and motion. • Middle level: Examples include presence or arrangement of specific types of objects, roles and scenes. • High level: Include impressions, emotions and meaning associated with the combination of perceptual features. Examples include objects or scenes with emotional or religious significance.

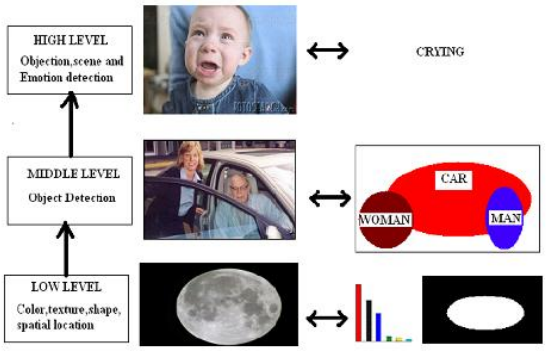


Fig. 1 Examples of Image Content Levels

The image content level is also a measure of level of feature extraction. At the low level, also regarded as primary level the features extracted (color, shape, texture, spatial information and motion) are called primitive features because they can only be extracted by information obtained at the pixel level, that is pixel representation of the images.

The middle level features are features that can be extracted by collection of pixels that make up the image, while high level features goes beyond the collection of pixels. It identifies the impressions, meanings and emotions associated with the collection of pixels that make up the object.

2.2 Feature Extraction

Feature selection and extraction is the pre-processing step of Image Mining. Feature extraction is a special form of dimensionality reduction in pattern recognition and in image processing. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called features extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of

variables generally requires a large amount of memory and computation power or a classification algorithm which overfits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

2.3 Image Information Mining

Image Information Mining Image Information Mining (IIM), in analogy to other forms of data mining, such as text-based information mining or semantic web technologies, aims at making semantic image content accessible. It is being evaluated at Isferea for its capability to automatically highlight satellite image subsets which contain specific objects of interest, such as settlements. If successful, it could assist in damage assessment by focusing time-consuming visual interpretation efforts on these relevant image areas. Additionally it is evaluated for its analysis capabilities on multi-temporal and multi-sensor datasets. In order to assess these capabilities two investigations have been conducted. Both are based on VHR optical satellite data used for operational damage assessment at the individual building level.

The first investigation concerns data fusion of multi-source VHR data in the IIM system KIM (Knowledge-Based Information Mining). In this framework guidelines have been developed for a data pre-processing and analysis sequence suitable for both visual interpretation and IIM-based data analysis in a rapid mapping scenario.

The second investigation assesses KIM's capability to reliably highlight image subsets based on specific semantic content as a component in an operational rapid mapping workflow. This included an evaluation of the system's probabilistic labeling accuracy as well as the reliability of its query functionality.

2.4 Wavelets

When digital images are to be viewed or processed at multiple resolutions, the discrete wavelet transform is the mathematical tool of choice. A wavelet is a wave-like oscillation with an amplitude that starts out at zero, increases, and then decreases back to zero. It can typically be visualized as a "brief oscillation" like one might see recorded by a seismograph or heart monitor. Generally, wavelets are purposefully crafted to have specific properties that make them useful for signal processing. Wavelets can be combined, using a "shift, multiply and sum" technique called convolution, with portions of an unknown signal to extract information from the unknown signal.

2.5 Independent Component Analysis

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed non gaussian and mutually independent, and they are called the independent components of the observed data. These independent components, also called sources or factors, can be found by ICA.

ICA is superficially related to principal component analysis and factor analysis. ICA is a much more powerful technique, however, capable of finding the underlying factors or sources when these classic methods fail completely. The data analyzed by ICA could originate from many different kinds of application fields, including digital images, document databases, economic indicators and psychometric measurements. Typical examples are mixtures of simultaneous speech signals that have been picked up by several microphones, brain waves recorded by multiple sensors, interfering radio signals arriving at a mobile phone, or parallel time series obtained from some industrial process.

3. METHODOLOGY

3.1 Coarse Image Segmentation

3.1.1 Improving Feature Set via ICA-Wavelet Transformation

Suitable transformation of the observed data is an important step for feature extraction and image segmentation. Previous work which converts the RGB space to the HSV space to make color component perceptually independent and uniform .However , the components in the HSV space are not statistically uncorrelated. Another disadvantage is that it uses only three components at a time, thus limiting its ability to capture the complete spectral pattern from the multispectral image. Hence, this work introduces the use of a statistical method ICA for data transformation along the spectral axis of the geospatial data, which transforms observed multivariate data into spectral components that are statistically independent and uncorrelated from each other.

3.1.2. Clustering

Two clustering algorithms, K-means and Kernel Kmeans, are considered on this study.

a) Clustering via the k-means Algorithm

This work uses the k-means algorithm for unsupervised segmentation. This algorithm provides relative scalability and very efficient processing for very large datasets compared to the hierarchical clustering algorithm. It is an iterative algorithm based on the squared-error function criterion.

b) Clustering via a kernel k-means Algorithm

Linear separation of the cluster can be obtained via the kmeans algorithm. Hence, during the feature extraction stage, cross-correlation energies, which actually represent few features from the higher dimensional space, are also extracted. A kernel approach transforms the feature set into a high dimensional feature space by smooth and continuous nonlinear mapping to provide a non-linear separation of the clusters data points. Hence, using the kernel approach in clustering eliminates the need to use the cross-correlation energies as features.

3.2. Region Recognition

3.2.1. Region Feature Formation

Research has shown that "energy-based" feature extraction from a region using the wavelet coefficients is more robust to image alterations such as object rotation, and high frequency sub-bands are very effective and robust for discriminating textures. The logarithmic localized energy obtained after applying a smoothing filter K (Gaussian filter) to a region is a good combination to describe a region. Thus the energy feature for each region in a sub-band is obtained using.

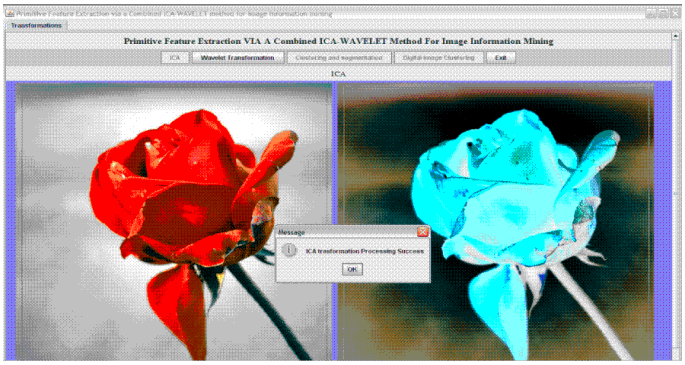
3.2.2. Fusing Missing Texture Features from a Panchromatic Image

Remote sensing data may consist of high-resolution panchromatic (PAN) images that have additional spatial characteristics. Thus, in this work, the missing textural information is obtained from the panchromatic bands. For example, for LandSat 7 ETM+ images, the resolution of the multispectral images is 30m, while the resolution of a PAN image is 15m. The localized energy for this high frequency sub-bands would be obtained using equation 1. Thus, if r is the ratio of the resolution of the multispectral image to the PAN image, then the total number of additional features would be 3 * (r -1) and the complete feature vector of the region after the fusion of feature vectors would be P + (3 * (r-1)) –dimension.

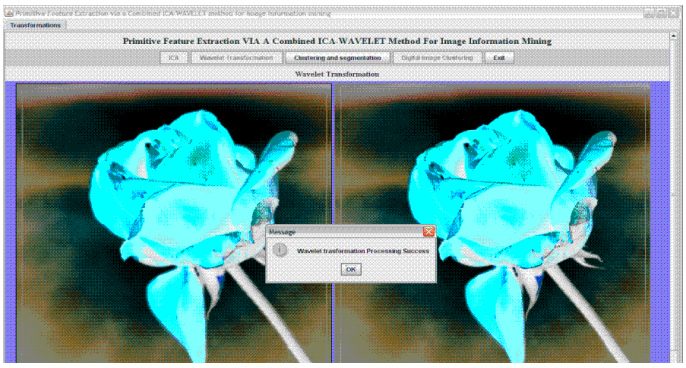
3.2.3. Support Vector Machines (SVM) for Region Identification

This work uses a new-generation learning system, Support Vector Machines (SVMs), for the identification of the region in an image. The SVMs have been successfully used for solving non-linear classification and regression problems. A few of the applications using the SVMs for classification are handwritten digit recognition, object detection, and face recognition. For example, a RBF kernel has a parameter σ (standard deviation), which must be selected a priori.

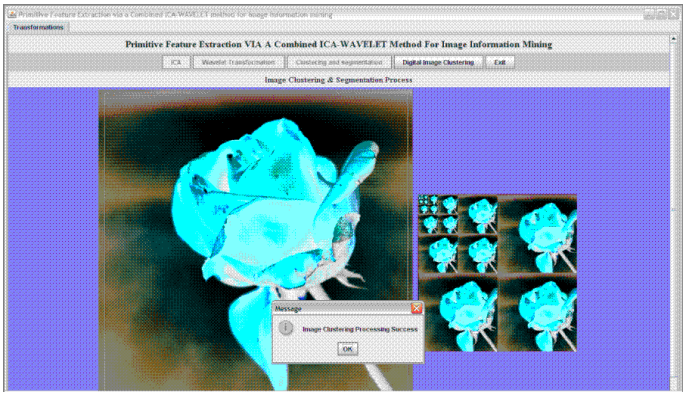
4. RESULT AND ANALYSIS



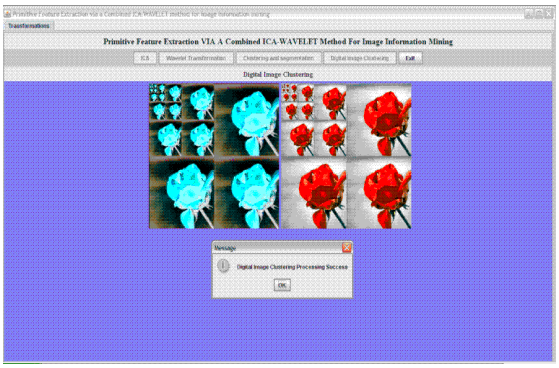
ICA Transformation (Spectral)



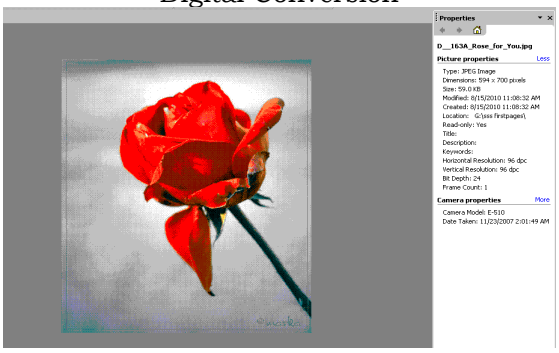
Wavelet Transformation(Spatial)



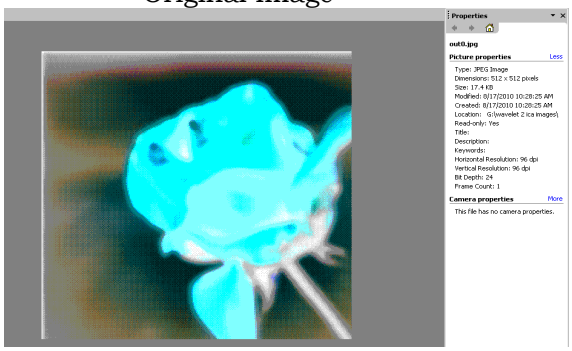
Clustering and Segmentation Process



Digital Conversion



Original Image



Transformed Image(Wavelet to ICA)

Moment	Expression	Measure of Texture
Mean	$m = \sum_{i=0}^{L-1} z_i \, p(z_i)$	A measure of average intensity.
Standard deviation	$\sigma = \sqrt{\mu_2(z)} = \sqrt{\sigma^2}$	A measure of average contrast.
Smoothness	$R = 1 - 1 / (1 + \sigma^2)$	Measures the relative smoothness of the intensity in a region. R is 0 for a region of constant intensity and approaches 1 for regions with large excursions in the values of its intensity levels. In practice the variance used in this measure is normalized to the range[0,1] by dividing it by $(L-1)^2$.
Third moment	$\mu_3 = \sum_{i=0}^{L-1} (z_i - m)^3 \, p(z_i)$	Measures the skewness of a histogram. This measure is 0 for symmetric histograms, positive by histograms skewed to the right(about the mean) and negative for histograms skewed to the left
Uniformity	$U = \sum_{i=0}^{L-1} p^2(z_i)$	Measures uniformity .This measure is maximum when all gray levels are equal(maximally uniform) and decreases from there.
Entropy	$e = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$	A measure of randomness.

4.2 Statistical Texture Measures

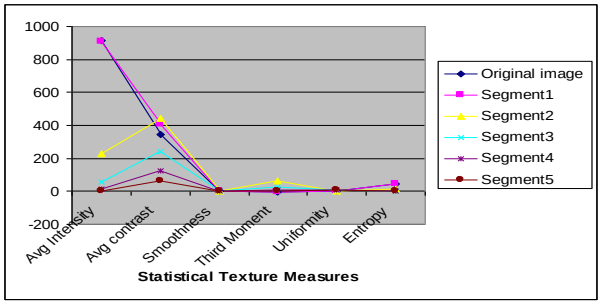
An important approach for describing a region is to quantify its texture content. A frequently used approach for texture analysis is based on statistical properties of the intensity histogram. One class of such measures is based on statistical moments. Descriptors of texture based on the intensity histogram of a region

4.3 Statistical Texture Measure Values of Various Regions

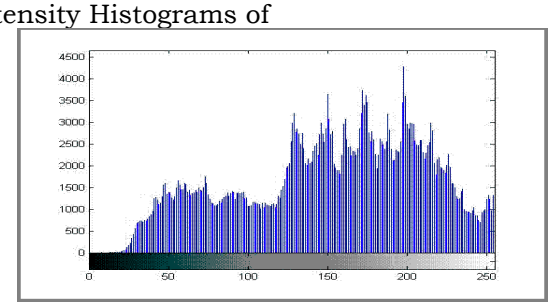
The following Table 1 shows the statistical texture measure values of various regions.

Regions	Avg Intensity	Avg contrast	Smoothness	Third Moment	Uniformity	Entropy
Original image	914.0762	346.8607	0.2933	-6.1593	0.0309	46.3117
Segment1	905.4994	405.5576	0.3939	-6.7829	0.0302	46.4236
Segment2	225.8168	440.1502	0.4586	63.2458	3.3636	16.5934
Segment3	56.3103	239.9434	0.1440	26.4560	5.2651	4.9861
Segment4	14.0393	121.8047	0.0378	7.2691	5.8091	1.4545
Segment5	3.5020	60.7284	0.0094	1.8122	5.9509	0.4173

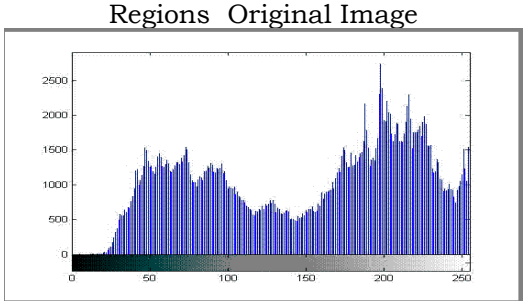
The Following figure shows the Statistical Texture Properties of Various Regions.



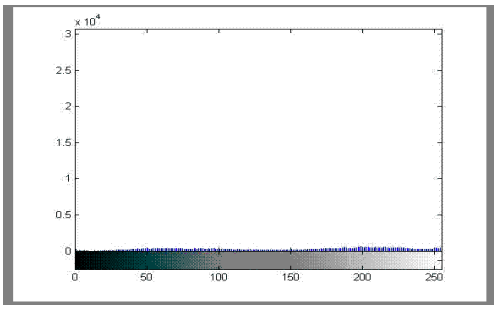
4.4 Statistical Properties of a Texture Based on Intensity Histograms of



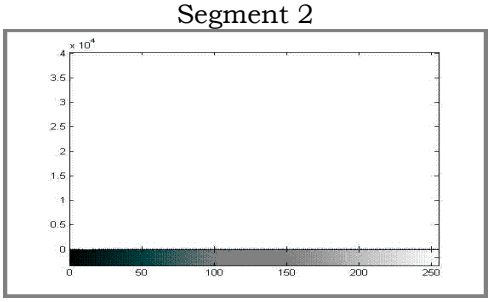
Regions Original Image



Segment1



Segment 2



Segment3

5. CONCLUSION:

This paper introduces the concept of using features obtained via a combined ICA-wavelet transformation. Experimental results suggest that the new feature sets are more effective in comparison to traditional approach in capturing both spectral and texture information for image mining. Feature reduction along the spectral dimension could be performed during the pre-whitening stage of calculating the ICA components.. Overall the system provides satisfactory results in retrieving the images for different classes used in this study. The presented approach helps in reducing the computation complexity for feature extraction process. Features obtained by ICA transformation provide reliable segmentation compared to other transformation approaches. Choice of the order between the spectral and spatial transform can quantitatively affect the image segmentation results. For the ICA-spectral transformation, the estimated number of clusters in an image mostly remains the same.

6. REFERENCES

[1] S. S. Durbha and R. L. King, "Semantics-enabled framework for knowledge discovery from Earth observation data archives," IEEE Transactions on Geosciences and Remote Sensing, Volume 43, Issue 11,Nov. 2005 pp. 2563 – 2572.

[2] V.P. Shah, N. H. Younan, S. S. Durbha, and R. L. King, " Wavelet Features for Information Mining in Remote Sensing Imagery," IEEE Int'l Conf. Geoscience and Remote Sensing Symposium, vol. 8, pp. 5630 – 5633, July 2005.

[3] S. Ardizzoni, I. Bartolini, and M. Patella, "Wind-surf: Region-based image retrieval using wavelets," in Proceedings of the 10th Int'l Workshop on Database & Expert Systems Applications, 1999.

[4] L. De Lathauwer, B. De Moor, J. Vandewalle, "An Introduction to Independent Component Analysis", Journal of Chemometrics, Special Issue on Multi-Way Analysis, vol. 14, no. 3, May-Jun. 2000, pp. 123-149.

[5] M. Bressan, D. Guillamet, J. Vitria, "Using an ICA representation of high dimensional data for object recognition and classification," Pattern Recognition, vol. 36, pp. 691-701, 2003.

[6] J. Shawe-Taylor and N. Cristianini, "Kernel Methods for Pattern Analysis," 1st edition, Cambridge University Press, 2004.

[7] M. Girolami, "Mercer Kernel-Based Clustering in Feature Space," IEEE Trans. of Neural Networks, vol. 13, no. 3, pp. 780-784, May 2002.

[8] B. Scholkopf, A. Smola, and K. R. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," Neural Comput., vol. 10, no. 5, pp. 1299-1219, 1998.

[9] V. N. Vapnik, Statistical Learning Theory. New York: Wiley, 1998.

[10] Yongqing Sun and Shinji Ozawa, "Efficient Wavelet-Based Image Retrieval Using Coarse Segmentation and Fine Region Feature Extraction," IEICE Trans. Inf. & Syst. Vol. E88-D, No. 5 May 2005.