Contrast Enhancement and Clustering Segmentation of Gray Level Images with Quantitative Information Evaluation

ZHENGMAO YE[1], HABIB MOHAMADIAN[1], SU-SENG PANG[2], SITHARAMA IYENGAR[2]
[1]Southern University, Baton Rouge, LA 70813, USA
zhengmaoye@engr.subr.edu, mohamad@engr.subr.edu
[2]Louisiana State University, Baton Rouge, Louisiana 70803
mepang@me.lsu.edu, iyengar@csc.lsu.edu

Abstract– Improper illumination and medium dispersing could occur in quite some gray level image collecting processes. Contrast enhancement and clustering segmentation are two effective approaches for the related pattern recognition problems. Image enhancement and image segmentation can be applied to different areas of science and engineering, such as biometric identification, national defense and resource exploration. Thus, adaptive image enhancement can be implemented to improve the image quality and to reduce random noise simultaneously, which is used to adapt to the intensity distribution within an image. Nonlinear K-means clustering can be applied for image segmentation, which is to classify an image into parts which have strong correlations with objects in order to reflect the actual information being collected. For example, it can be used against the effects from unevenly distributed pressure or temperature condition under atmosphere medium and water medium. To evaluate the actual roles of image enhancement and image segmentation, some quantity measures should be taken into account. In this study, a set of quantitative measures is proposed to evaluate the information flow of gray level image processing. Concepts of the gray level energy, discrete entropy, relative entropy and mutual information are proposed to measure outcomes of the adaptive image enhancement and K-means image clustering.

Key-Words: - Contrast Enhancement, K-Means Segmentation, Gray Level Image, Energy, Entropy, Relative Entropy, Mutual Information

1 Introduction

Image enhancement and image segmentation can be applied to different areas of science and engineering. Except for illumination conditions, quality of images is also affected by external noises and environmental disturbances such as ambient pressure and temperature fluctuations. Thus, image enhancement is necessary. Approaches of contrast limited image enhancement via stretching the histograms over a reasonable dynamic range and multi-scale adaptive histogram equalizations can be developed. An adaptive algorithm is adapted to the image intensity distribution either globally or locally. By separating smooth and detail areas of an image, the algorithm is applied to each of them to avoid excessive enhancement of noises. In most cases, quality of images is affected by atmosphere medium and water medium, therefore image segmentation is required, which is to classify an image into parts which have strong correlations with objects in order to reflect the actual information. Optimization is often associated with image clustering, where the winner-take-all (WTA) algorithm, nearest neighbor rule and self organization network can be applied [3-7].

To measure effects of image processing, information measures need to be introduced. Stem from thermo-dynamics, concepts of entropy, relative entropy, energy and mutual information can be expanded to a number of areas in signal processing and image processing [1-2]. From literatures, some interesting results have been documented. For instance, the entropy of a fuzzy set is viewed as a global measure of the fuzziness of the fuzzy set and the energy of a fuzzy set is viewed as a local measure, where the trade relationship exists between the entropy of a fuzzy set and information energy. In another case, the concept of relative entropy is treated as the fitness function of genetic algorithms for the segmentation thresholding methods. As an unsupervised clustering technique, fuzzy C-Means algorithm has been used to retrieve from the database the color JPEG images with more accurate results. Different geometric shape descriptors are presented for evaluating color image enhancement and color image segmentation based upon the study of the equivalent class set to color vectors. Additionally, in comparison with the potential discrete entropy from Principle of Maximum Entropy, enhancement for several different types of images can be investigated.

These measures are selected in practice to quantify image enhancement, like image compression [8-13]. Accordingly, these measures are proposed to evaluate effects of locally adaptive image enhancement techniques in this research. On a basis of gray level histograms, measures of the gray level energy, discrete entropy, relative entropy and mutual information of images processes can be used to indicate information flow throughout the image processing process [14-18].

2 Adaptive Histogram Equalization for Image Enhancement

Image contrast enhancement is achieved using the histogram equalization algorithms. It is inevitable that certain types of noises could be amplified or induced simultaneously. For this reason, the adaptive image contrast enhancement scheme based on histogram equalization is conducted. The images are separated into a number of small regions. Within every small region, the histogram is obtained under contrast limit and then the exponential distribution is applied as a basis to create the contrast transform function. The new gray level assignment of pixels in each small local region is specified by this contrast transformation for image enhancement. The gray level mapping from each local histogram is then generated. In order to avoid the occurrence of boundary artifacts arisen from different neighborhoods of small regions and to obtain an evenly distributed overall enhancement across the entire regions, the bilinear interpolations among four different mappings are used together so that contrast saturation problem is solved via contrast limit, until results from adaptive equalization are satisfactory. Through this simple approach, excessive enhancement of noises can be minimized.

As a general approach, this method has been applied to a group of gray scale images, namely, the scenery image of mount Hollywood, US capitol architecture image, static southernmost point image and dynamic deer image on summit of the Grouse Mountain. For these gray scale images, histogram contains 256 bins and the percentage of counts for each bin over the summation will give rise to its probability distribution. The parameter for exponential functions, the factor for contrast limiting and weights for bilinear interpolation may be adjusted to achieve quality outcomes. Using this adaptive image enhancement methodology, from Fig. 1 to Fig. 4, two types of original and enhanced scenery and architecture images are listed; From Fig. 5 to Fig. 8, another two types of original and enhanced static object and living creature images are listed.

Fig. 1 Original Image of Mount Hollywood

Fig. 2 Enhanced Image of Mount Hollywood

Fig. 3 Original Image of US Capitol

Fig. 4 Enhanced Image of US Capitol

Fig. 5 Original Image of US Southernmost Point

Fig. 6 Enhanced Image of US Southernmost Point

Fig. 7 Original Deer Image on Grouse Mountain

Fig. 8 Enhanced Deer Image on Grouse Mountain

3 Nonliear K-Means Clustering for Image Segmentation

Image segmentation is a useful working approach to identify object contents for data analysis. It requires the number of clusters be specified for partitioning. Centers of each cluster are defined initially, which represent mean values of all those data points in that cluster. A distance metric must be defined to quantify the relative distances of objects. The Euclidean and Mahalanobis distances are 2 typical cases of distance metrics. Distance measure is based on the spatial grayscale color histograms of original underwater or aerial images. The Mahalanobis metric distance has been used, which is formulated as:

$$d=(s-X_A)^T K_A^{-1} (s- X_A) \quad (1)$$

where X_A is the cluster center of any layer A, s is any point, d is the Mahalanobis distance, K_A^{-1} is inverse of covariance matrix.

K-means clustering assigns each object a space location. It classifies data sets through K number of clusters. An optimal statistical algorithm is usually selected for classification, which can be threshold based, region based, edge based or surface based. The distances of any data point to different cluster centers are compared. The winner-take-all (WTA) competitive learning technique is used in k-means algorithms so that only one cluster centroid is updated for each input. Images will be decomposed into three recognized physical entities. Clustering depends on partition of images into a set of layers. The WTA learning network classifies input vectors into one of three specified numbers of categories according to clusters detected in the training dataset. The learning is performed in an unsupervised mode. Each cluster center has its own associated weight that is referred to as w_i. The winner has been defined as one whose cluster center is closest to inputs. Thus, this mechanism allows the competition for all input responses, but only one output is active each time. The unit that wins the competition is the WTA cluster.

$w_{ij}x = \min(w_i x)$ for j = 1, 2, 3 and i = 1, 2, 3 (2)

$w_{i1} + w_{i2} + w_{i3} = 1$ for i = 1, 2, 3

Assume cluster center Y wins. Then the weight increment of only Y is computed and updated as (3), where α is a small positive learning constant and α decreases as the competitive learning progresses.

$\Delta w_{ij} = \alpha(x_j - w_{ij})$, for j = 1, 2, 3 and i = 1, 2, 3 (3)

To minimize the distortion, the K-means clustering algorithm iterates between the cluster point labeling and cluster center reassigning. Point labeling groups data points belong to the same cluster. At each iteration, points are reassigned to the winning center. Center reassigning recalculates centers for all clusters. This algorithm iterates between point labeling and center reassigning procedures until the convergence to local stationary is achieved.

Fig. 9 Yellow Sea Original Image

All input data are sent to these algorithms, which will identify all three clusters and associated samples with the selected winning cluster. The criterion corresponds to finalizing the weight vector that is closest to the input. Only winner weight vector is adjusted. With a proper initial input vector, the cluster center converges representing final clusters properly after enough number of iterations. Via clustering procedures, objects are close to each other within each of three clusters, and far from objects of other clusters. The winner is actually the one with the weight vector closest to the input vector. In Figs 10-12, three clusters of aerial images are shown.

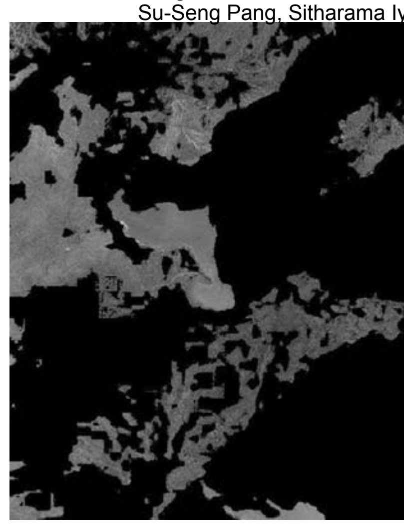

Fig. 10 Cluster 1 of Yellow Sea Image

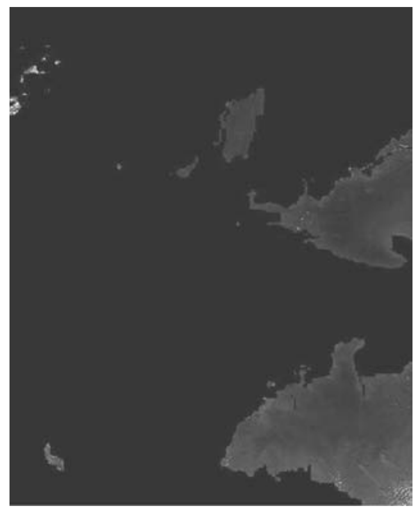

Fig. 11 Cluster 2 of Yellow Sea Image

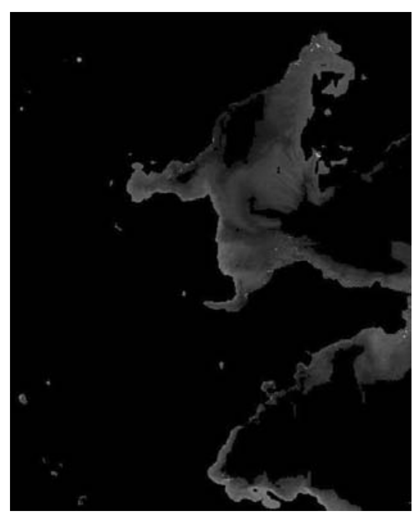

Fig. 12 Cluster 3 of Yellow Sea Image

4 Histogram and Probability Function

The histogram plot is used to display the brightness of the gray scale image, showing the occurrence of pixel counts for all 256 intensity levels. Occurrence of the gray level configuration may be described as co-occurrence matrices of relative frequencies. Then the classification is based on features being derived from co-occurrence matrices of the images. The occurrence probability function of the gray level can be simply estimated from the histogram, which is formulated in (4), where p(k) is the probability distribution function and h(k) is the histogram function.

$$p(k)=\frac{h(k)}{\sum h(k)} \tag{4}$$

In Figs. 13-18, probability distributions of different original and processed images are plotted, which are used for calculation of quantitative measures.

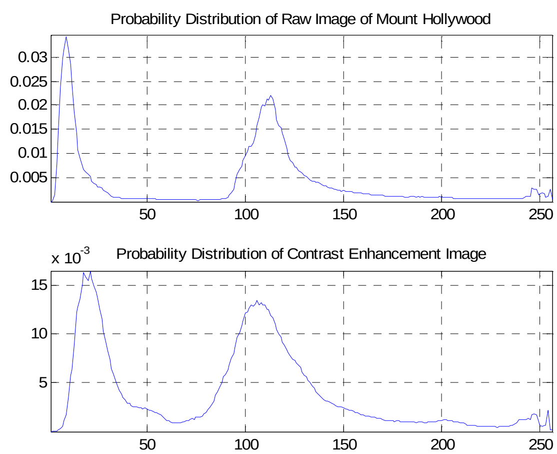

Fig. 13 Probability Distribution of Mount Hollywood Image

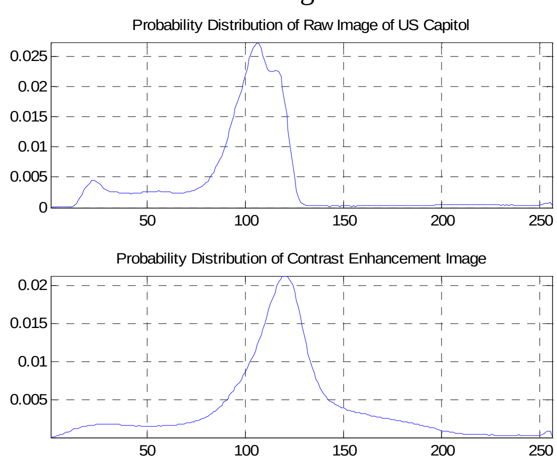

Fig. 14 Probability Distribution of US Capitol Image

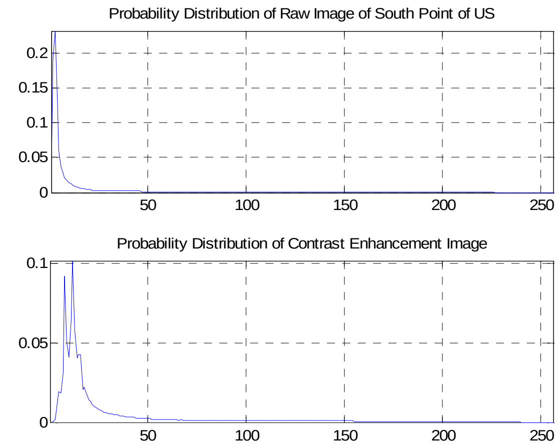

Fig. 15 Probability Distribution of US South Point

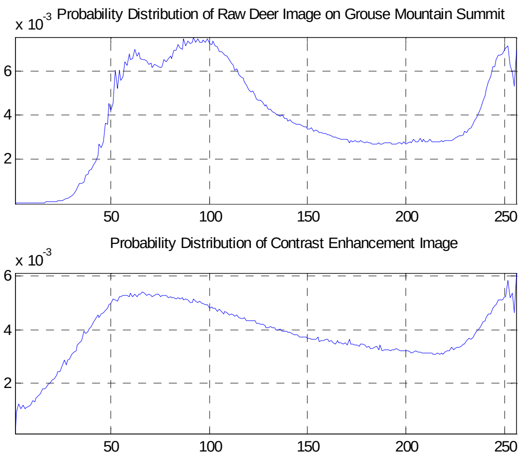

Fig. 16 Probability Distribution of Deer Image

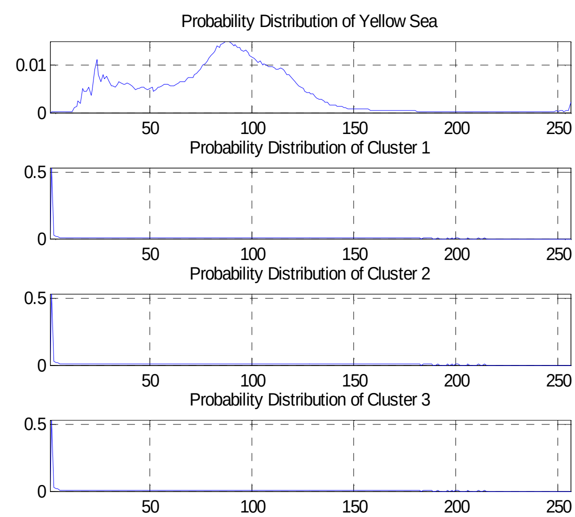

Fig. 17 Probability Distribution of Image Clustering

5 Discrete Gray Level Energy

The gray level energy measure indicates how the gray levels are distributed. The formulation of the gray level energy is shown in (5), where E(x) represents the gray level energy with 256 bins and p(i) refers to the probability distribution functions under different gray levels, which contains histogram counts. For a special case when an image has a constant value of gray level, the energy measure reaches its maxima of 1, as shown in (6). An image with larger energy can be compressed much easier than one with the smaller energy. The larger energy corresponds to a lower number of gray levels (simple) and the smaller one corresponds to a higher number of gray levels (complex).

$$E(x)=\sum_{i=1}^{k} p(i)^2 \qquad (5)$$

$$\max\{E(x)\}=\max\{\sum_{i=1}^{k} p(i)^2\}=1 \qquad (6)$$

Table 1. Energy in Image Processing

Original Image	Energy	Enhanced Image	Energy
Hollywood	0.0144	Hollywood	0.0086
South Point	0.1185	South	0.0400
US Capitol	0.0052	US Capitol	0.0042
Mount Deer	0.0161	Mount Deer	0.0107
Raw Image	Cluster 1	Cluster 2	Cluster 3
0.0089	0.2911	0.2895	0.2929

In Table 1, it can be seen that the enhanced images using the adaptive histogram equalization algorithms have the smaller gray level energy to some extent, which means enhanced images are relatively complex for all cases whose individual probability functions decrease through nonlinear transformation and bilinear interpolation. For K-means clustering, all 3 individual clusters are much simpler than the raw image.

6 Discrete Entropy

Entropy is the measure of the image information content, which is interpreted as the average uncertainty of the information source. Discrete entropy is the summation of the products of the probability of outcome multiplied by the log of the inverse of probability of outcome, taking into considerations of all possible outcomes {1, 2, …, n} in the event $\{x_1, x_2, \ldots, x_n\}$, where n is the gray level; p(i) is the probability at the gray level of i, which contains all the histogram counts. It is formulated as (7-8).

$$H(x)=\sum_{i=1}^{k} p(i)\log_2\frac{1}{p(i)} = -\sum_{i=1}^{k} p(i)\log_2 p(i) \qquad (7)$$

$$\sum_{i=1}^{k} p(i)=1 \qquad (8)$$

The discrete entropy is a statistical measure of randomness. The maximal entropy occurs when the probabilities of all potential outcomes are equal. When the outcome is a certainty, minimal entropy occurs which is equivalent to zero. For image processing, the discrete entropy is a measure how many bits needed for coding image data. Discrete entropy of different original images and processed images are shown in Table 2. It actually represents an average amount of information conveyed from each image. The results have shown that the entropy of the enhanced image is slightly higher using adaptive histogram equalization algorithms and bilinear interpolation. From another aspect, it shows that enhanced images are relatively complex. A relatively complex image has higher entropy than a relatively simple image. When the pixels in the image are distributed among more gray levels, values of the corresponding discrete entropy increase. Similarly for K-means clustering, individual clusters are simpler than the raw image, thus smaller values are obtained than that of the raw image.

Table 2. Entropy in Image Processing

Original Image	Entropy	Enhanced Image	Entropy
Hollywood	6.7456	Hollywood	7.2698
South Point	4.3094	South	5.7259
US Capitol	7.7024	US Capitol	7.9365
Mount Deer	6.4975	Mount Deer	7.1024
Raw Image	Cluster 1	Cluster 2	Cluster 3
7.0377	4.0337	4.0265	4.0418

Principle of maximum entropy can also be applied to analyze the potential of image enhancement and image compression. Assume mutually exclusive propositions have individual discrete probability distributions. The minimum information entropy of an image is equal to zero when one of the distributions is definitely true, representing the most informative distribution case. On the other hand, when the distribution is uniform, the maximum discrete entropy occurs with the discrete entropy value of $\log_2(n)$ = 8 bits (n=256). In this case, no proposition is superior to other propositions, thus, the least informative distribution occurs. The discrete entropy provides a numerical measure between zero and $\log_2(n)$, from most informative case to the totally uninformative case. All quantities of discrete entropy in these examples are within a range between 0 and 8, the latter of which is the maximum entropy possible.

7 Relative Entropy

Suppose two discrete probability distributions of the processing images have the probability functions of p and q. Relative entropy of p with respect to q is then defined as the summation of all possible states of the system, which is formulated as (9).

$$d=\sum_{i=1}^{k} p(i)\log_2 \frac{p(i)}{q(i)} \qquad (9)$$

Table 3. Relative Entropy in Image Processing

Enhanced v.s. Raw	Relative Entropy	Enhanced v.s. Raw	Relative Entropy
Mount Hollywood	0.6898	US Capitol	0.2567
US South Point	1.7372	Deer on Summit	1.1539
Raw v.s. Enhanced	Relative Entropy	Raw v.s. Enhanced	Relative Entropy
Mount Hollywood	*0. 9860*	*US Capitol*	*0.1255*
US South Point	*3.7373*	*Deer on Summit*	*0.6132*

Relative entropy is also referred to as the Kullback Leibler distance. In Table 3, relative entropy of enhanced images with respect to original images has been indicated. On the other hand, relative entropy of original images with respect to enhanced images can also be calculated, where p(i) and q(i) can be swapped in (9). The different set of results is listed in italic style. It is shown relative entropy is non-symmetric. The relative entropy of three clusters with respect to the raw image is listed in Table 4.

Table 4. Relative Entropy in Image Processing

Mutual Information	Cluster 1	Cluster 2	Cluster 3
Raw Image	1.8073	1.8113	1.8028

8 Mutual Information

From Table 3, relative entropy between a pair of original and enhanced images is different. Thus, relative entropy issues can be taken into account from both aspects. Another concept of mutual information I(X; Y) can be applied as well, which describes how much information one variable tells about the other variable. Its relationship is formulated as (10)

$$I(X;Y)=\sum_{X,Y} p_{XY}(X,Y)\log_2 \frac{p_{XY}(X,Y)}{p_X(X)p_Y(Y)}$$
$$= -\sum_{X} p_X(X)\log_2 p_X(X) + \sum_{X,Y} p_{XY}(X,Y)\log_2 \frac{p_{XY}(X,Y)}{p_Y(Y)} \qquad (10)$$
$$=H(X) - H(X|Y)$$

where H(X) and H(X|Y) are values of the entropy and conditional entropy; p_{XY} is the joint probability density function; p_X and p_Y are marginal probability density functions. It can be explained as information that Y can tell about X is the reduction in uncertainty of X due to the existence of Y. It can also be regarded as the relative entropy between joint distribution and product distribution. The results of mutual information in contrast enhancement and clustering segmentation cases are shown in Table 5.

Table 5. Mutual Information Between Images

Enhanced and Raw	Mount Hollywood	Southern Point	US Capitol	Mount Deer
Mutual Information	0.5239	1.4137	0.6102	0.2303

Mutual Information	Cluster 1	Cluster 2	Cluster 3
Raw Image	3.0040	3.0152	2.9914

From Table 1 to Table 4, it has been shown that the gray level energy, discrete entropy, relative entropy and mutual information have been successfully served as quantitative measures to evaluate the quality of image enhancement and image segmentation. The measures are useful to shorten the decision making process in pattern recognition problems.

9 Conclusions

The adaptive contrast enhancement methodology has been applied to processing of various types of images using histogram equalization algorithms. Then useful information is extracted and enhanced to avoid feature ambiguity. To eliminate artifacts generated by noises throughout the image enhancement, the local adaptive histogram equalization is applied which is followed by the interpolations among neighborhoods in order to eliminate artificial boundaries. The image contrast has been stretched and results are satisfactory. Nonlinear K-means clustering has also been applied for image segmentation, which is to classify an image into parts which have strong correlations with objects in order to indicate intrinsic information. The optimal approach is selected for clustering. This approach also applies to all possible small and large scale image processing. The results have shown the effectiveness of this approach. Several quantitative measures are used to evaluate the image processing techniques. The gray level energy, discrete entropy, relative entropy and mutual information are all computed through the image processing approaches. It provides valuable studies for a wide variety of image processing fields.

References

[1] T. M. Cover and J. A. Thomas, "Elements of information theory", 2nd Edition, John Wiley and Sons, 2005

[2] D. MacKay, "Information Theory, Inference and Learning Algorithms", Cambridge Press, 2003

[3] R. Duda, P. Hart and D. Stork, “Pattern Classification,” 2nd Edition, Wiley & Sons, 2000

[4] R. Gonzalez and R. Woods, “Digital Image Processing,” 2nd Edition, Prentice-Hall, 2002

[5] S. Hykin, “Neural Networks - A Comprehensive Foundation”, 2nd Edition, 1999, Prentice Hall

[6] J. Stark, “Adaptive Image Contrast Enhancement using Generalization of Histogram Equalizations”, IEEE Transactions on Image Processing 9(5): 889-896, May, 2000

[7] Z. Ye, Y. Ye, H. Mohamadian, P. Bhattacharya and K. Kang, "Fuzzy Filtering and Fuzzy K-Means Clustering on Biomedical Sample Characterization", 2005 IEEE International Conference on Control Applications, pp. 90-95, Aug. 28-31, 2005, Toronto, Canada

[8] T. Ji, M. Sundareshan, H. Roehrig, “Adaptive image contrast enhancement based on human visual properties”, IEEE Transactions on Medical Imaging, Vol. 13, Issue 4, pp. 573 - 586, 1994

[9] I. Vlachos, G. Sergiadis, "Image quality indices based on fuzzy discrimination information measures", WSEAS Transactions on Circuits and Systems, v 3, n 2, pp. 241-6, April 2004

[10] D. Niyomua, S. Supratid and C. Kimpan, "Improving image retrieval by fuzzy C-means initialized by fixed threshold clustering: Case studies relating to a color temperature histogram and a color histogram", WSEAS Transactions on Systems, v 5, n 9, pp. 2182-2188, September, 2006

[11] Z. Ye, H. Mohamadian and Y. Ye, "Discrete Entropy and Relative Entropy Study on Nonlinear Clustering of Underwater and Arial Images", Proceedings of the 2007 IEEE International Conference on Control Applications, pp. 318-323, Oct. 1-3, 2007, Singapore

[12] B. Bhanu, S. Lee and S. Das, “Adaptive Image Segmentation Using Genetic and Hybrid search methods,” IEEE Trans. on Aerospace and Electronic systems, Vol. 31, No. 4, Oct. 1995

[13] Y. Ye, Z. Ye, J. Luo, P. Bhattacharya, H. Majlesein, R. Smith, "On Linear and Nonlinear Processing of Underwater, Ground, Aerial and Satellite Images", 2005 IEEE International Conference on Systems, Man and Cybernetics, pp. 3364-3368, Oct. 10-12, 2005, Hawaii, USA

[14] P. Refregier and F. Goudail, “Kullback Relative Entropy and Characterization of Partially Polarizes Optical Waves”, Journal of the Optical Society of America A, Vol. 23, No.3, pp. 671-8, March, 2006

[15] Z. Ye, H. Mohamadian and Y. Ye, "Information Measures for Biometric Identification via 2D Discrete Wavelet Transform", 2007 IEEE International Conference on Automation Science and Engineering, pp. 835-840, Sept. 22-25, 2007, Scottsdale, Arizona, USA

[16] J. Wang, Y. Du, C. Chang; “Relative Entropy-Based Methods for Image Thresholding”, 2002 IEEE International Symposium on Circuits and Systems, May 26-9, 2002, Phoenix, USA

[17] Z. Ye, J. Luo and P. Bhattacharya, "Optimal Approach for Enhancement of Large and Small Scale Near-Infrared and Infrared Imaging", 2005 IEEE International Conference on Industrial Electronics and Control Applications (ICIECA05), 6 pp., Nov. 29 - Dec. 2, 2005, Quito, Ecuador

[18] A. Fotinos, G. Economou and E. Zigouris, “The Use of Entropy for Color Edge Detection”, 6th IEEE International Conference on Electronics, Circuits and Systems, Vol.1, pp. 437-40, 1999