Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

An Efficient Image Pattern Recognition System Using an Evolutionary Search Strategy

Pei-Fang Guo, Prabir Bhattacharya and Nawwaf Kharma

***Abstract*—A mechanism involving evolutionary genetic programming (GP) and the expectation maximization algorithm (EM) is proposed to generate feature functions automatically, based on the primitive features, for an image pattern recognition system on the diagnosis of the disease OPMD. Prior to the feature function generation, we introduce a novel technique of the primitive texture feature extraction, which deals with non-uniform images, from the histogram region of interest by thresholds (HROIT). Compared with the performance achieved by support vector machine (SVM) using the whole primitive texture features, the GP-EM methodology, as a whole, achieves a better performance of 90.20% recognition rate on diagnosis, while projecting the hyperspace of the primitive features onto the space of a single generated feature.**

***Keywords*—evolutionary computation, pattern recognition, feature generation, image processing, genetic programming, the expectation maximization algorithm, Gaussian mixture estimation, texture analysis, moments**

I. Introduction

Feature generation is generally considered a process of mapping the primitive features into more effective features during the run of the algorithms. Many researchers experimented with various problems for the techniques of feature generation, such as artificial neural network (ANNs) for machine condition monitoring devices [1], fuzzy systems for health monitoring systems [2], and evolutionary algorithms (EA) for object recognition [3] and the classification of machinery faults [4].

Inspired by biological evolution, EA is a generic population-based optimization algorithm [5]. Evolutionary genetic programming (GP), introduced by Koza [6], has been widely employed in various applications. A methodology for online feature select (FS) and classifier design was proposed using GP in [7]. A progression of approaches involving GP was described in the problem of the landmark detection [8].

Generally, the expectation maximization algorithm (EM) can be applied in many settings where we wish to estimate some set of parameters $\boldsymbol{\theta}$ that describe an underlying probability distribution [9]. The EM algorithm has been employed in missing value prediction [10] and the signal detection problem in wireless communication systems [11].

Texture analysis is based on statistical properties of the intensity histogram. Histograms have been the most popular strategy for intensity level computation owing to their simplicity and efficiency [12]. However, for images of different sizes and shapes with localized background structures, the use of the simple whole regions of the channel histogram for the image pattern classification would influence the classification accuracy.

One way to solve this problem would be to restrict the bin length to a region that captures the knowledge of the images within that particular part regardless of the sizes, shapes and backgrounds.

This study is the first to present a generalized primitive texture feature extraction technique based on a histogram region of interest by thresholds (HROIT), a bin-length-based technique, for characterizing diagnostically the disease OPMD. Further to extend the idea of an automatic image pattern recognition system, a new technique of the feature generation, composed of the primitive features extracted from HROIT, is proposed by means of integrating GP and the EM algorithm (GP-EM).

This paper is organized as follows. Section II introduces the primitive feature preparation. In Section III, we summarize the proposed GP-EM system. The EM algorithm, from which the *k*-means problem is derived, is presented in Section IV. The GP application is described in Section V. Experimental results are given in Section VI, and the final section is the conclusion.

II. Primitive Feature Preparation

The use of the feature generation requires primitive features as a prerequisite for the subsequent applications. In this study, the first phase is to prepare our experimental primitive feature set which consists of three steps (1) image grayscale conversion; (2) HROIT definition; and (3) statistical texture feature extraction. The procedure is illustrated in Fig. 1, using the database called CellsDB.

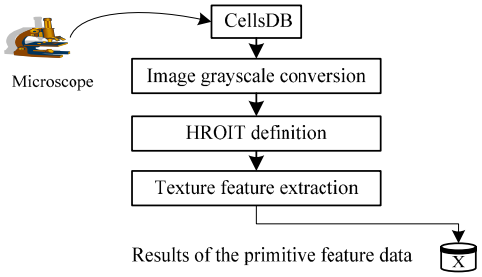

Figure 1. The procedure of the primitive feature extraction.

Recorded from microscopic imagery, CellsDB is a database of pre-segmented cell images divided into two categories,

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

healthy and sick, associated with Oculopharyngeal Muscular Dystrophy (OPMD) disease. Each of the images in CellsDB, a non-uniform image group, contains a single cell on a white background at an optimal level of RGB values. The presence of inclusions in patients' cell images leads to the suggestion that protein aggregation is a critical molecular component of such OPMD disease. Our conducted research requires the identification of the inclusions within the cell images. The samples, a healthy cell and a sick cell, from CellsDB are shown in Fig. 2.

(a) A typical healthy cell image (b) A typical sick cell image

Figure 2. Sample images from CellsDB.

A. Cell Image Grayscale Conversion

Conversion of a color image to grayscale is very common in digital image processing [12]. To be able to capture the inclusion of color information, we begin our color analysis by converting the cell images into grayscale. The reasons to choose image grayscale conversion are: (1) It reduces three 2D R, G and B histograms to an independent grayscale image without losing any information; (2) It filtrates the cell image background in the margin of the histogram regardless of differences in size and shape.

The resultant number of the grayscale cell images is the desired linear luminance value. Each of the grayscale images is commonly stored with 8 bits per sampled pixel, which allows 256 different intensities (i.e., shades of gray) to be recorded.

B. HROIT Definition

When reading the cell images, the difference between the healthy and sick ones is whether the inclusions are inside the cells or not. Normally, the healthy cell images are dark green, regardless of their sizes and shapes. The inclusions within the sick cells with a bumpy appearance, which indicates the disease OPMD, are often a lighter green color corresponding to the middle region of the grayscale histogram.

If taking into consideration the whole regions of the grayscale histogram, the pixel intensity can be mis-computed because of differences in size and shape of the cell images. From this miscomputation, we could miss the inclusion intensity information in terms of their discriminant capabilities, i.e., their sick versus healthy.

In this case, where the lighter inclusions cover the middle region of the grayscale histogram, we use two thresholds to limit an interval of gray values, and then define this interval as the histogram region of interest by thresholds (HROIT). Thereby, HROIT is able to capture only the inclusion intensity information within the cell images and not the cell structural differences.

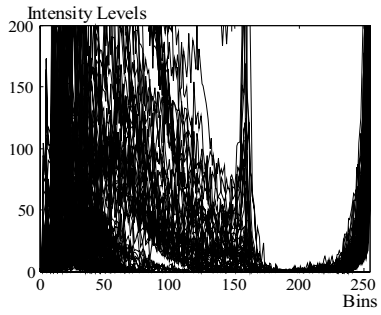

Figure 3. Grayscale histograms of all image samples.

As illustrated in Fig. 3, the upper threshold, T_{upper}, is set at the bin of 200 by analyzing the output of all the image grayscale histograms because there is a zero intensity region centered at the value of this bin.

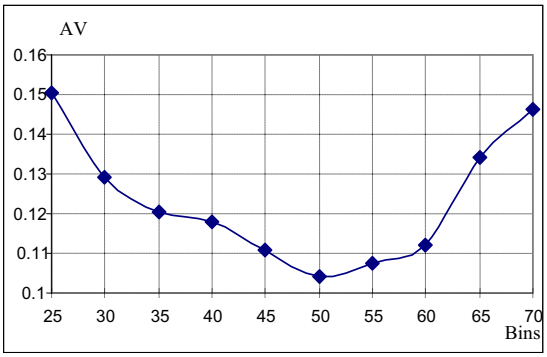

Figure 4. The lower threshold, T_{lower}, at the bin of 50, defined by the average of variance (AV).

Since pixel intensities for both healthy and sick images overlap in the lower intensity boundary on the bins ranged (25, 70), we use the variance to determine the lower threshold, T_{lower}. In statistics, the variance is one measure of statistical dispersion, averaging the squared distance of its possible values from the expected value (the mean). The smaller values of the variance represent that the features have relatively lower variability to the mean [13].

In Fig. 4, the average of the variances (AV) measures the scale or degree of being spread out of the primitive texture feature set in the bin range of (25, 70). A feature set with good stability and reliability should exhibit a small AV among the samples. The optimal value of the bin should be at the lowest AV, so we set the lower threshold, T_{lower}, at the bin of 50.

Fig. 5 illustrates the design of HROIT limited to a bin range of (50, 200), with the grayscale histograms, which are converted from two typical cell images, healthy and sick, shown in Fig. 2.

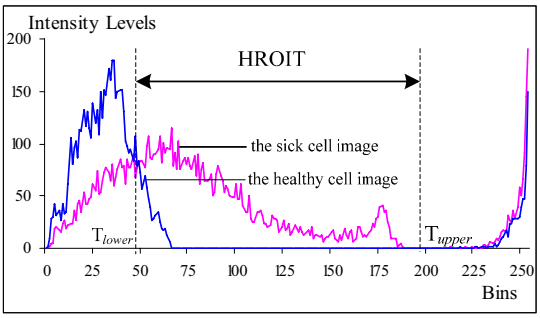

Figure 5. HROIT representation, with the grayscale histograms of two samples in Fig. 2.

C. Statistic Texture Features Extracted from HROIT

To characterize the complexity of the CellsDB images, the primitive feature extraction primitive feature relates to their textural characteristics such as randomness, uniformity, etc. Additionally, the advantage of moments is straightforward, and moments also carry a 'physical' interpretation of boundary shape. The texture features extracted from HROIT are as follows:

- *Entropy* (*F1*): measures the randomness or homogeneity of the histogram distribution with respect to length or orientation.

$$F1 = -\sum_{}^{L_{HROIT}} p(z_i)\log_2 p(z_i)$$

- *Angular Second Momentum* (*F2*): measures the uniformity.

$$F2 = \sum_{}^{L_{HROIT}} p^2(z_i)$$

- *Mean* (*F3*): measures the average of intensity.

$$F3 = \sum_{}^{L_{HROIT}} z_i p(z_i)$$

- *Moments* (*F4~ F8*): measure the five (second to sixth order) moments.

$$Fl = \sum_{}^{L_{HROIT}} z_i^n p(z_i) \quad n = 2,3,4,5,6 \quad and \quad l = 4,5,6,7,8$$

- *Peak Density* (*F9*): measures the strength of the local dominant peak of the intensity.

$$F9 = \max_{}^{L_{HROIT}}\{p(z_i)\}$$

- *Range of the Density* (*F10*): measures the range of the intensity.

$$F10 = range_{}^{L_{HROIT}}\{p(z_i)\}$$

- *Median Density* (*F11*): measures the median value of the intensity.

$$F11 = median_{}^{L_{HROIT}}\{p(z_i)\}$$

where z_i indicates intensity, $p(z_i)$ is the histogram of the intensity levels, and L_{HROIT} is the number of intensity levels within HROIT.

At the end, a total of 11 statistic texture features are normalized to center it at zero mean and scale it to unit standard deviation.

III. The GP-EM Recognition System

As a result of the primitive feature data **X** (the 11-dimensional feature data) extracted from HROIT, the approach to the feature generation for the given task of the diagnosis of the disease OPMD could be employed. Fig. 6 illustrates the evolutionary process of the GP-EM system.

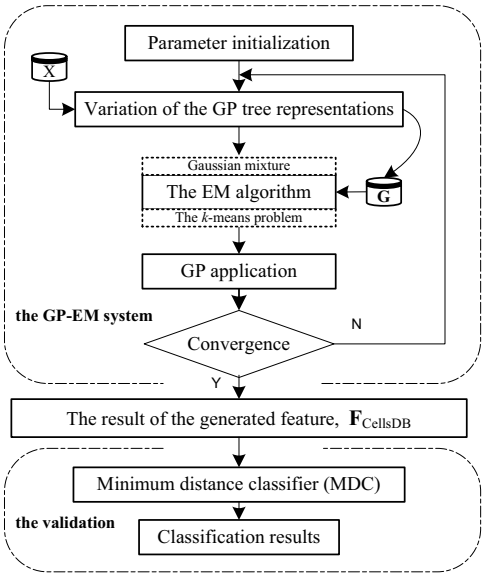

Figure 6. The GP-EM recognition system for the CellsDB images.

Suppose we measure the primitive feature data **X** with *m* samples and the *k* known classes. The generated feature functions, based on the GP tree representations, are going to provide the data mapping from the primitive feature data **X** to the generated values, $g_i \in \mathbf{G}$.

The parametric EM algorithm assumes that its input mapped data **G** follows a particular probability Gaussian distribution for each class for estimation of the parameters vector $\boldsymbol{\theta} = \{\boldsymbol{\theta}_j = (\boldsymbol{\pi}_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), j = 1, \ldots, k\}$.

Since each of the *k* Gaussian distributions has the same variance with uniform probability chosen by each independently drawn Gaussian signal g_i, we get the univariate case that the mean vector, $\boldsymbol{\mu}_j$, becomes the unknown parameter vector only. The derivative of the *k*-means problem via the EM algorithm is described in Section IV.

Then, under the hypothesis of the *k*-means problem via the EM, the performance of the GP-EM system is going through a

typical GP process (described in Section V), including the selection of genetic operators and the fitness measure.

At the end of the evolutionary GP-EM training process, the system returns the best tree as the input to the minimum distance classifier (MDC) in the validation stage.

IV. THE *K*-MEANS PROBLEM VIA THE EM ALGORITHM

Given the data **G**, a set of *m* generated values mapped from the data **X**, the PDF of **G**, which follows *k* component finite mixture distributions f_j evaluated at g_i, can be written as [11]:

$$p(g_i / \boldsymbol{\theta}) = \sum_{j=1}^{k} \boldsymbol{\pi}_j f_j(g_i / \boldsymbol{\theta}_j) \qquad j = 1,..., k \tag{1}$$

where $\boldsymbol{\pi}_j$ is the mixing probability.

The component density functions $f_j(g_i/\boldsymbol{\theta})$ with a *d*-dimensional mean vector $\boldsymbol{\mu}_j$ and a $d \times d$ covariance matrix $\boldsymbol{\Sigma}_j$ is:

$$f_j(g_i / \boldsymbol{\theta}_j) = \frac{\exp(-\frac{1}{2}(g_i - \boldsymbol{\mu}_j)^T (\boldsymbol{\Sigma}_j)^{-1}(g_i - \boldsymbol{\mu}_j))}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_j|}} \tag{2}$$

where the parameters $\boldsymbol{\theta} = \{(\boldsymbol{\pi}_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j), j = 1,\ldots, k\}$ are determined to measure how well the corresponding mixture model fits the given data **G**.

One of the most common methods is to assume equal prior probabilities for all categories by setting $P(\boldsymbol{\pi}_j) = 1 / k$ in which *k* is the number of classes [14]; all the groups have the same covariance matrix and terms involving this constant can be cancelled, which means the covariance matrix $\boldsymbol{\Sigma}_j$ contains only one component [15]: the variance σ^2. Therefore, the parameters $\boldsymbol{\theta} = \{(\boldsymbol{\mu}_j), j = 1,\ldots, k\}$ is the complete set that describes the means of each of the *k* distributions.

The log-likelihood of the probability of the complete data **Y** for a mixture of the *k* Gaussian distributions is [16]:

$$\ln P(\mathbf{Y} / h') = \sum_{i=1}^{m} (\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2} \sum_{j=1}^{k} z_{ij} (g_i - \boldsymbol{\mu}_j')^2) \tag{3}$$

where the revised hypothesis $h' = (\mu_1',\ldots,\mu_k')$ and z_{ij} is a binary *k* dimensional vector that has the value *1* if g_i was created by the *j*th distribution and *0* otherwise.

After taking the expected value of $\ln P(\mathbf{Y} / h')$ and applying the maximum likelihood (ML) estimator, the EM algorithm repeats the following steps in the estimation of the *k*-means problem [16] [9]:

$$E[z_{ij}] = \frac{p(g = g_i / \boldsymbol{\mu} = \boldsymbol{\mu}_j)}{\sum_{n=1}^{k} p(g = g_i / \boldsymbol{\mu} = \boldsymbol{\mu}_n)} = \frac{\exp(-\frac{1}{2\sigma^2}(g_i - \boldsymbol{\mu}_j)^2)}{\sum_{n=1}^{k} \exp(-\frac{1}{2\sigma^2}(g_i - \boldsymbol{\mu}_n)^2)} \tag{4}$$

Then, using $E[z_{ij}]$, the probability that instance g_i is generated by the *j*th Gaussian distribution (*E*-step), to obtain a new maximum likelihood hypothesis (*M*-step):

$$\boldsymbol{\mu}_j \leftarrow \sum_{i=1}^{m} E[z_{ij}] g_i \Big/ \sum_{i=1}^{m} E[z_{ij}] \tag{5}$$

Equations (4) and (5) define the two steps in the implementation of the GP-EM recognition system for the devising of the mixture model of the *k* Gaussians, in practice.

V. GP APPLICATION

One of the key features of the GP application is that it uses tree structure representations to solve problems. There are two major tasks processed in the GP loop: fitness measure and genetic operators to allow for the identification of the best individuals in a probabilistic way; see a detailed description of GP in [6].

A. The Function Set and the Terminal Set

Table I lists the function set, a mathematics elementary function set, that attempts to fit the given problem in this study. Division and square root operations are protected against divisions by zero and negative values, respectively.

TABLE I. DESIGNS OF THE FUNCTION SET

Signs	Functions	Terminals
+, -	Addition, Subtraction	2
*, /	Multiplication, Division	2
S, C, T	Sine, Cosine, Tan functions	1
P, R, E	Square, Square root, Exponential	1
A, N	Absolute, Negative value	1

For the purpose of this experiment, the terminal set is associated with each of 11 texture features described in Section II and with an additional constant, *c*, involved.

B. GP Tree Representations

Fig. 7 shows a tree representing the feature function of 0.5 + (F1 – F3) x cos (F2). On the tree, functions are internal nodes, which represent elementary functions such as addition and multiplication, etc; and terminals are leaf nodes which receive the primitive features.

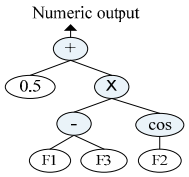

Figure 7. A GP tree representation: 0.5 + (F1 - F3) x cos (F2).

C. Operator Selection

GP allocates every individual some chance of being selected to participate in the operations of reproduction, crossover and mutation [7].

The reproduction operator selects the best individuals on the basis of their fitness by copying it into the new population.

Crossover operator is performed by replacing a randomly chosen sub-tree of one parent program by a sub-tree from the other parent program. Fig. 8 illustrates the crossover operation.

The basic idea of the GP mutation for terminal and function operators is that any noninvariant point anywhere in the overall program is randomly chosen, without restriction.

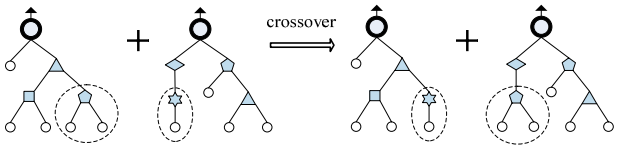

Figure 8. GP sub-tree crossover.

D. Fitness Function

Suppose there is a set of m instances in the data **G** divided into the k subsets $\mathbf{G}_1,\ldots,\mathbf{G}_k$. The within-class scatter matrix is [16]:

$$\boldsymbol{S}_W = \sum_{j=1}^{k} \sum_{g \in \boldsymbol{G}_j} (g - \boldsymbol{\mu}_j)(g - \boldsymbol{\mu}_j)^T \qquad (6)$$

then the between-class scatter matrix is:

$$\boldsymbol{S}_B = \sum_{j=1}^{k} n_j (\boldsymbol{\mu}_j - \boldsymbol{\mu}_T)(\boldsymbol{\mu}_j - \boldsymbol{\mu}_T)^T \qquad (7)$$

where $\boldsymbol{\mu}_j$ is replaced by the mean of the jth Gaussian distribution, $\boldsymbol{\mu}_T$ is the total mean vector, $\mathbf{G}_j$ is the jth subset of **G** and n_j is the number of samples in the jth class.

In this study, we choose an optimal partition measurement as the fitness function by using the logarithm of J value which equals to the ratio of the between-class matrix and the within-class scatter matrix:

$$J = \ln |\boldsymbol{S}_W^{-1} \boldsymbol{S}_B| \qquad (8)$$

VI. Experimental Results

The experiments use 500 cell images, 250 healthy and 250 sick, randomly selected from CellsDB. We choose medium population size and relatively large number of generations to help the algorithm adequately sample the search surface to find better solutions of the generated features. Table II lists the GP parameters used for the CellsDB experiments.

TABLE II. GP Parameters

Parameter Names	Values
Population Size	32
No. of Generations	400
Max. of Tree Levels	5
No. of Operations in the Function Set	12
Reproduction Probability, P_r	0.125
Crossover Probability, P_c	0.50
Mutation Probability, P_m	0.375

For classification, we employ the classifier MDC by running the 10-fold cross-validation method, the statistical practice to randomly divide the data into mutually exclusive partitions (folds) for training and testing.

A. Convergence Results

In the GP-EM system, we stop the GP for 400 consecutive generations when repeated applications of genetic operators yield little effort on the fitness associated with the most elite solution in the current. The fitness results, normalized in the ranged of (0, 1), for a single experiment are shown in Fig. 9.

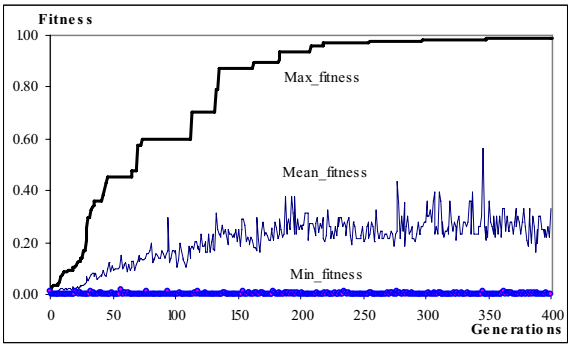

Figure 9. Convergence results.

On an average, the time for each of the 10 independent runs took about four minutes on a Pentium 4 PC, with CPU 1GHz, 1GB of RAM and an 80 mega-byte hard disk drive. Microsoft Visual C++ 6.0 was used for implementing the GP-EM algorithm.

B. Feature Generated Results

After convergence, the resulting generated feature, the best, of a LISP-like expression for the single experiment is:

$\mathbf{F}_{\text{CellsDB}}$ = (+(*(+(+F10F10–F9F9)*(S(*F6F2)T(/F4F7)))
((+F4F12 +F1F2) C(/(*F12F6 E(/F7F5))))))

where F_l (l = 1,…,11) refers to 11 primitive texture features described in Section II, and F12 is a random constant which equals 0.572027; the signs of '+', '-', '*' , '/', 'S', 'C', 'T' and 'E' represent *sine*, *cosine*, *tan* and *exponential* functions, respectively, which are listed in Table I.

A rational expression of the resulting generated feature is:

$$\mathbf{F}_{\text{CellsDB}} = 2\text{F10} \times \sin(\text{F6} \times \text{F2}) \times \tan(\text{F4}/\text{F7}) + (\text{F4} + 0.572027) \times (\text{F1} + \text{F2}) \times \cos(0.572027 \times \text{F6}/e^{(\text{F7}/\text{F5})})$$

C. Classification Results and Comparisons

1) Classification Results for 10 independent runs

For the cell image data, the statistical measure for 10 independent runs on the combined training sets and test sets is illustrated in Table III. We achieve the average training score of 91.40% accuracy. With only one generated feature, $\mathbf{F}_{\text{CellDB}}$, the average performance over all 10 runs shows that 90.20% of the test data can be classified correctly.

TABLE III. The Results of Classification Accuracy (CA) for 10 Independent Runs on CellsDB

Statistics	Training sets (CA)	Test sets (CA)
Maximum (%)	96.00	92.00
Minimum (%)	88.00	82.00
Average (%)	91.40	90.20
Std (%)	3.16	3.50

2) Comparisons of the GP-EM system and the SVM

To evaluate the proposed GP-EM recognition system and demonstrate its advantages, the experiments on a linear support vector machine (SVM) classifier are carried out with comparisons.

The basic idea of the SVM is to construct a hyperplane as the decision plane, which separates the positive (+1) and negative (-1) classes with the largest margin, the sum of the distances from the hyperplane to the closest data points of each of the two classes [17].

Table IV illustrates the comparison results of classification recognition rates between the GP-EM system with one generated feature, $\mathbf{F}_{\text{CellDB}}$, and the classifier SVM with 11 primitive texture features over 10 independent runs for the CellsDB images. Significantly, for the test CA, the GP-EM system achieves the average performance of 90.20%, better than the performance achieved by the SVM approach.

TABLE IV. COMPARISON RESULTS OF CLASSIFICATION ACCURACY (CA) BETWEEN THE GP-EM SYSTEM AND THE SVM APPROACH FOR 10 INDEPENDENT RUNS ON CELLSDB

Statistics	The GP-EM system using one generated feature (CA)	The classifier SVM using 11 primitive texture features (CA)
Maximum (%)	92.00	81.00
Minimum (%)	82.00	69.00
Average (%)	**90.20**	78.60
Std (%)	3.50	4.10

VII. CONCLUSION

In this study, we propose a new fully automatic GP-EM recognition system for classification applied to the problem of the diagnosis of the disease OPMD. Via the EM algorithm, the GP-EM generated data is modeled as a Gaussian mixture, in which the learning task results in a simpler hypothesis of the *k*-means problem. On the other hand, the evolutionary GP process is driven by a fitness measure accounting for inter-class separability and intra-class homogeneity. Overall, the proposed GP-EM system using only one generated feature leads to higher classification accuracies with a lower standard deviation, compared with the performance of the SVM approach using 11 primitive features.

In addition to the main contribution, there are other unique and novel attributes and tools provided by this study for biomedical image processing:

(1) The proposed GP-EM recognition system not only shows the capability of enhancing the classification performance but also succeeds in reducing the dimensionality from the hyperspace of the primitive features to a 1-dimensional $\mathbf{F}_{\text{CellsDB}}$-space.

(2) The structures of the biomedical images affect the quality of the primitive feature extraction, a prerequisite for the learning task of the feature generation. The proposed HROIT, an innovative histogram-based technique, shows the applicability to non-uniform images.

(3) With the filtration of the image background in the margin and the application of the thresholds, the resultant of image grayscale histogram is able to encompass all the inclusion information contained in the histograms of three color channels, including those which cannot be captured by single intensities of 2D R, G and B histograms.

The current system has a limitation of the shortage of the terminators. This can be solved by implementing more primitive features. Therefore, the feature search space improvement and the algorithm development will be done in our future work.

ACKNOWLEDGMENT

This work is supported by grants from the NSERC and Canada Research Chair Foundation.

REFERENCES

[1] M. L. D. Wong, L.B. Jack, and A. K. Nandi, "Modified self-organising map for automated novelty detection applied to vibration signal monitoring," *Mechanical Syst. and Signal Process.*, vol.20, pp. 593-610, 2006.

[2] G. G. Yen, and P. Meesad, "An effective neuro-fuzzy paradigm for machinery condition health monitoring," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 31, pp.523–536, 2001.

[3] K. Krawiec, and B. Bhanu, "Visual learning by coevolutionary feature synthesis," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, pp.405-425, 2005.

[4] P. Chen, T. Toyota, and Z. He, "Automated function generation of symptom parameters and application to fault diagnosis of machinery under variable operating conditions," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol.31, pp.775-781, 2001.

[5] A. E. Eiben, and J. E. Smith, Introduction to Evolutionary Computing, NY: Springer, 2003.

[6] J. R. Koza, Genetic Programming II: Automatic Discovery of Reusable Programs, Cambridge, MA: MIT Pr., 1994.

[7] D. P. Muni, N. R. Pal, and J. Das, "Genetic programming for simultaneous feature selection and classifier design," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol.36, 106-117, 2006.

[8] V. Ciesielski, G. Wijesinghe, A. Innes, and S. John, "Analysis of the superiority of parameter optimization over genetic programming for a difficult object detection problem," *IEEE Congress on Evol. Comput.*, pp.1264-1271, 2006.

[9] T. M. Mitchell, Machine Learning, NY: McGraw-Hill, 1997.

[10] Lin, T. I., Lee, J. C., and Ho, H. J. On fast supervised learning for normal mixture models with missing information. *Pattern Recog.*, 39 (2006), 1177-1187.

[11] F. Chan, and J. Choi, "Neighborhood exploring detector: An EM-based signal detector for multiple antenna systems," *IEEE Trans. Signal Process,* vol.55, pp.1875-1885, 2007.

[12] R. C. Rafael, Digital Image Processing, MA: A. Wesley, 2002.

[13] G. Keller, and B. Warrack, Statistics for Management and Economics, NY: Duxbury Press, 2000.

[14] M.L. Raymer, T.E. Doom, L.A. Kuhn, and W.F. Punch, "Knowledge discovery in medical and biological datasets using a hybrid bayes classifier/evolutionary algorithm raymer," *IEEE Trans Syst., Man, Cybern. B, Cybern.*, vol. 33, pp. 802-813, 2003.

[15] M. James, Classification Algorithms, London: Collins, 1985.

[16] C. M. Bishop, Pattern Recognition and Machine Learning, NY: Springer, 2006.

[17] S.W. Lee, and A. Verri, Pattern Recognition with Support Vector Machines (LNCS 2388), NY: Springer-Verlag, 2002.