Automation of DNA Computing Readout Method Implemented on LightCycler System

[1]Muhammad Faiz Mohamed Saaid, [1]Zuwairie Ibrahim, [1]Marzuki Khalid, and [2]Nor Haniza Sarmin
[1]*Center for Artificial Intelligence and Robotics (CAIRO)*
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
[2]*Department of Mathematics*
Faculty of Science
Universiti Teknologi Malaysia
ninjagaban@gmail.com, zuwairie@fke.utm.my, marzuki.khalid@gmail.com, nhs@mel.fs.utm.my

Abstract

In this paper, an automation of DNA computing readout method based on real-time Polymerase Chan Reaction (PCR) is developed, which employs a hybrid in vitro-in silico approach. In the in vitro phase, TaqMan-based real-time PCR reactions are performed in parallel, to investigate the ordering of pairs of nodes in the Hamiltonian Path Problem (HPP), in terms of relative distance from the DNA sequence encoding the known start node. The real-time PCR experiment is implemented on LightCycler system. Fuzzy C-Means (FCM) clustering algorithm is used to identify automatically two different reactions in real-time PCR, followed by in silico algorithm, which in turn, enables extraction of the Hamiltonian path. A software called SILICOLIGHT is built to implement the FCM clustering and the in silico algorithm, which return the desired Hamiltonian path.

1. Introduction

The subsequent innovation of real-time PCR has rapidly gained popularity and plays a crucial role in molecular medicine and clinical diagnostics [1]. All real-time amplification instruments require a fluorescence reporter molecule for detection and quantitation, whose signal increase is proportional to the amount of amplified product.

A TaqMan DNA probe is a modified, nonextendable duallabeled oligonucleotides. The 5' and 3' ends of the oligonucleotide are terminated with an attached reporter, such as FAM, and quencher fluorophores dyes, such as TAMRA, respectively, as shown in Figure 1 [2]. Upon laser excitation at 488 nm, the FAM fluorophore, in isolation emits fluorescence at 518 nm. Given proximity of the TAMRA quencher, however, based on the principle of fluorescence resonance energy transfer (FRET), the excitation energy is not emitted by the FAM fluorophore, but rather is transferred to TAMRA via the dipole-dipole interaction between FAM and TAMRA. As TAMRA emits this absorbed energy at significantly wavelengths (580 nm), the resulting fluorescence is not observable in Channel 1 of real-time PCR instruments [3].

Previously, we proposed a readout method tailored specifically to the HPP in DNA computing, which employs a hybrid *in vitro-in silico* approach [4]. In the in vitro phase, $O(|V|^2)$ TaqMan-based real-time PCR reactions are performed in parallel, to investigate the ordering of pairs of nodes in the Hamiltonian path of a $|V|$-node instance graph, in terms of relative distance from the DNA sequence encoding the known start node. The resulting relative orderings are then processed *in silico*, which efficiently returns the complete Hamiltonian path. The proposed approach is experimentally validated optical method specifically designed for the quick readout of HPP instances, in DNA computing. Previously, graduated PCR, which was originally demonstrated by Adleman [5], was employed to perform such operations. While a DNA chip based methodology, which makes use of biochip hybridization for the same purpose has been proposed [6-7], this method is more costly, and has yet to be experimentally implemented.

In this paper, an automation system for DNA computing readout method implemented on real-time PCR is developed, which consist of in vitro-in silico approach. The real-time amplification is performed with the TaqMan probes and the TaqMan detection mechanism was exploited for the design and development of the readout approach. The *in vitro* part

978-0-7695-3648-4/09 $25.00 © 2009 IEEE
DOI 10.1109/AMS.2009.80

Figure 1. Illustration of the structure of a TaqMan DNA probe. Here, R and Q denote the reporter and quencher fluorophores, respectively

is performed on the LightCycler System. As shown in Figure 2, the output of DNA computing readout method implemented on LightCycler System consist of two kinds of reactions, namely "YES" reaction and "NO" reaction. In the *in silico* information processing, Fuzzy C-Means (FCM) [8] clustering algorithm is implemented for automatic classification of "YES" and "NO" reaction. A software called *SILICOLIGHT* is developed and able to implement the in silico information processing during the *in silico* phase of readout approach in order to cluster the ''YES'' and ''NO'' reaction, followed by *in silico* algorithm which directly show the Hamiltonian path.

2. Notations and basic principle

First of all, $v_{1(a)}v_{2(b)}v_{3(c)}v_{4(d)}$ denotes a double-stranded DNA (dsDNA) which contains the base-pair subsequences, v_1, v_2, v_3, and v_4, respectively. Here, the subscripts in parenthesis (a, b, c, and d) indicate the length of each respective base-pair subsequence. For instance, $v_{1(a)}$ indicates that the length of the double-stranded subsequence, v_1 is 20 base-pairs (bp). When convenient, a dsDNA may also be represented without indicating the segment lengths (*e.g.*, $v_1v_2v_3v_4$).

A reaction denoted by TaqMan(v_0,v_k,v_l) indicates that real-time PCR is performed using forward primer v_0, reverse primer v_l , and TaqMan probe v_k . Based on the proposed approach, there are two possible reaction conditions regarding the relative locations of the TaqMan probe and reverse primer. In particular, the first condition occurs when the TaqMan probe specifically hybridizes to the template, between the forward and reverse primers, while the second occurs when the reverse primer hybridizes between the forward primer and the TaqMan probe. As shown in Figure 3, these two conditions would result in different amplification patterns during real-time PCR, given the same DNA template (*i.e.*, assuming that they occurred separately, in two different PCR reactions). The higher fluorescent output of the first condition is a typical amplification plot for real-time PCR. In contrast, the low fluorescent output of the second condition reflects the cleavage of a few of the TaqMan probes via DNA polymerase due to the 'unfavourable' hybridization position of the reverse primer. Thus, TaqMan(v_0,v_k,v_l) = YES if an amplification plot similar to the first condition is observed, while TaqMan(v_0,v_k,v_l) = NO if an amplification plot similar to the second condition is observed.

3. Real-time PCR based readout approach

Let the output of an *in vitro* computation of an HPP instance of the input graph be represented by a 120-bp dsDNA $v_{0(20)}v_{2(20)}v_{4(20)}v_{1(20)}v_{3(20)}v_{5(20)}$, where the Hamiltonian path $V_0 \rightarrow V_2 \rightarrow V_4 \rightarrow V_1 \rightarrow V_3 \rightarrow V_5$, begins at node V_0, ends at node V_5, and contains intermediate nodes V_2, V_4, V_1, and V_3, respectively. Note that in practice, only the identities of the starting and ending nodes, and the presence of all intermediate nodes will be known in advance to characterize a solving path. The specific order of the intermediate nodes within such a path is unknown.

The first part of the approach, which is performed *in vitro*, consists of $[(|V|\text{-}2)^2\text{-}(|V|\text{-}2)]/2$ real-time PCR reactions, each denoted by TaqMan(v_0,v_k,v_l) for all k and l, such that $0 < k < |V|\text{-}2$, $1 < l < |V|\text{-}1$, and $k < l$. For this example instance, so that the DNA template is the dsDNA, $v_0v_2v_4v_1v_3v_5$ these 6 reactions are as follows:

(1) TaqMan(v_0,v_1,v_2) = YES
(2) TaqMan(v_0,v_1,v_3) = YES
(3) TaqMan(v_0,v_1,v_4) = YES
(4) TaqMan(v_0,v_2,v_3) = YES
(5) TaqMan(v_0,v_2,v_4) = YES
(6) TaqMan(v_0,v_3,v_4) = NO

Note that the overall process consists of a set of parallel real-time PCR reactions, and thus requires $O(1)$ laboratory steps for *in vitro* amplification. The accompanying SPACE complexity, in terms of the required number of tubes is $O(|V|^2)$. Clearly, only one forward primer is required for all real-time PCR reactions, while the number of reverse primers and TaqMan probes required with respect to the size of input graph are each $|V|$-3.

After all real-time PCR reactions are completed, the *in vitro* output is subjected to an algorithm for *in silico* information processing, producing the satisfying Hamiltonian path of the HPP instance in $O(n^2)$ TIME (here, n denotes vertex number). The next step is to use all the information from the six TaqMan reactions to allocate each node of the Hamiltonian path. This can be done by applying the *in silico* algorithm as follows.

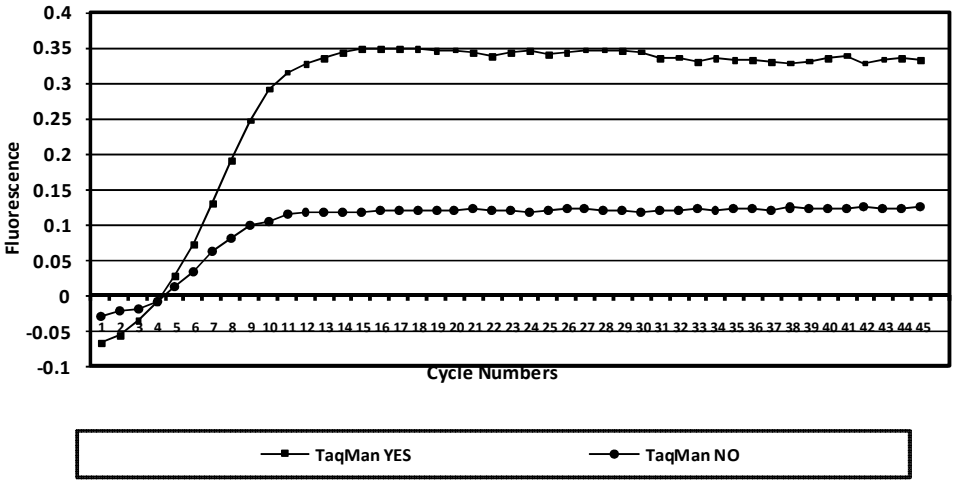

Figure 2. An example of reaction plots corresponding to TaqMan(v_0,v_k,v_l) = YES(first condition) and TaqMan(v_0,v_k,v_l) = NO (second condition).

Input: N[0…|*V*|-1]=2 // N[0, ?, ?, ?, ?, 5]
A[1…|*V*|-2]=|*V*| // A[1, 1, 1, 1]
  for *k*=1 to |*V*|-3
    for *l*=k+1 to |*V*|-2
      if TaqMan(v_0,v_k,v_l) = YES
        A[*l*] = A[*l*]+1
      else A[*k*] = A[*k*]+1
      endif
    endfor
    N[A[*k*]] = *k*
  endfor
N[A[|*V*|-2]]=|*V*|-2

In this algorithm, an array (N[0…|*V*|-1) that store all the nodes of the Hamiltonian path is defined. In addition, an array of aggregation values (A[1..|*V*|-2]) that is used to locate the Hamiltonian path in each array of nodes is also defined. Based on the modified algorithm, the input array N is first initialized to N={0,?,?,?,?,5} since the start and the end of the path are known, in advance. Next, the aggregation array A is initialized to A={1,1,1,1}. During the loop operations of the algorithm, the value of in the array A is increased in each iteration steps. The aggregation array A[*i*] is used for indexing the nodes array for each value of *k*. After the loop operation |*V*|-2 is assigned to the N[A[|*V*|-2]. The output of the *in silico* algorithm can be viewed by calling back all the nodes array N[0] to N[|*V*|-1]. The outcome of the this *in silico* algorithms is N={0,2,4,1,3,5}. Note that this algorithm can be done if all the information of TaqMan reactions is already achieved. This only can be done if clustering is applied to investigate "YES" and "NO" reaction.

4. Real-time PCR experiments

In this study, real-time PCR was performed on a LightCycler 2.0 Instrument (Roche Applied Science, Germany) where amplification is carried out in a 20 μl LightCycler capillary tube (Roche Applied Science, Germany). Two solutions were prepared for each reaction: (1) 3 μl of a 10x primer/probe solution (5 μM of primers and 1 μM of probes), prepared by mixing 0.75 μl of 20 μM forward primer solution, 0.75 μl of a 20 μM reverse primer solution, and 1.5 μl of a 2 μM probe solution; and, (2) 15 μl of PCR mix, containing 4 μl of reaction mix, 9 μl PCR grade/water, and 2 μl of the previous 10x primer/probe solution. Note that even though 3 μl of 10x primer/probe solution was prepared, only 2 μl of that solution was used for the preparation of PCR mix. 15 μl of PCR mix was then injected via pipette into a capillary tube. Afterwards, 5 μl of the input molecule solution was injected into the same capillary tube. The capillary tube was then sealed with a stopper, and placed in an adapter. The adapter containing the capillary was placed into a microcentrifuge, and the centrifugation was performed at 3000 rpm.

Six separate real-time PCR reactions were performed, in order to implement the first stage of the proposed HPP readout. The amplification consists of

45 cycles of denaturation, annealing, and extension, performed at 95ºC, 48ºC, and 72ºC, respectively. The annealing temperature is primer-dependent, and should be selected at 5ºC below the calculated primer melting temperatures. In the current study, the lowest primer melting temperature was estimated at 53.5ºC. Accordingly, an annealing temperature of 48ºC was selected. The resulting real-time PCR amplification plots are illustrated in Figure.3.

5. *In silico* information processing

5.1. Exporting data

The data from LightCycler System can be exported by clicking the export data on the software. From the data, only the FAM reading was selected to be exported into the *SILICOLIGHT* software, since the FAM emission reading is the actual result of TaqMan reaction shown in Figure 5. The exporting process was done after the in vitro phase is already completed. This exported data were written in the text file format, which show the fluorescence intensity from the first to the 45th thermal cycle for all 6 different TaqMan reactions.

5.2. Clustering algorithm

From the output graph of real-time PCR, an unsupervised learning such as clustering algorithm can be implemented for automatic classification of a data. Fuzzy C-means (FCM) has become the well known and powerful method in cluster analysis, and has been applied in many fields. FCM is a data clustering technique based on the optimization of the objective function [8]:

$$J(U,Y)=\sum_{i=1}^{C}\sum_{j=1}^{N}(\mu_{ij})^{m}\left\|x_j-y_i\right\|^2 \quad (1)$$

where C is a number of clusters and N is a number of data. Every data point in the data set requires to belong to a cluster at a particular membership degree. The purpose of FCM is to group data points into different specific clusters. Let $X=\{x_1, x_2,\ldots, x_N\}$ be a collection of data. By minimizing (1), X is classified into C homogeneous clusters, where μ_{ij} is the membership degree of data x_j to a fuzzy cluster set y_i, $Y=\{y_1, y_2,\ldots, y_C\}$ are the cluster centers. $U=(\mu_{ij})_{N*C}$ is a fuzzy partition matrix, and μ_{ij} indicates the membership degree of each data point in the data set to the cluster i. The value of U should satisfy the following conditions:

$$\mu_{ij}\in[0,1],\quad \forall i=1,\ldots C,\ \forall j=1\ldots N \quad (2)$$

$$\sum_{i=1}^{C}\mu_{ij}=1,\ \forall j=1,\ldots N \quad (3)$$

The $\|x_j - y_i\|$ is the Euclidean distance between x_j and y_i. The parameter m is called fuzziness value index, which control the fuzziness value of membership of each datum. The cluster center can be calculated by using the following equation:

$$y_i=\frac{\sum_{j=1}^{N}(\mu_{ij})^{m}x_j}{\sum_{j=1}^{N}(\mu_{ij})^{m}},\ \forall i=1,\ldots, C \quad (4)$$

Then, clustering can be achieved by iteratively minimize the aggregate distance between each data point in the data set and cluster centers until no further minimization is possible. Then, the fuzzy partition matrix U is updated by using following equation:

$$\mu_{ij}=\frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_j-y_i\right\|}{\left\|x_j-y_k\right\|}\right)^{\frac{2}{m-1}}} \quad (5)$$

In order to cluster the results of TaqMan reaction, namely "YES" and "NO" reaction, each graph of the reactions are represented as vector $\boldsymbol{x}_j=\{x_{j(1)},x_{j(2)},x_{j(3)},\ldots,x_{j(44)},x_{j(45)}\}$. The reactions are clustered into two groups, having their centre at $\boldsymbol{y}_1=\{y_{1(1)},y_{1(2)},\ldots,y_{1(44)},y_{1(45)}\}$ and $\boldsymbol{y}_2=\{y_{2(1)},y_{2(2)},\ldots,y_{2(45)},y_{2(45)}\}$. The two centres can be viewed as graphs that similar to the TaqMan reaction "YES" and "NO". It is noticed that the centre that is located in the amplification region always have greater value than the other center in the non-amplification region. The two centers are called as "YES" and "NO" center, where "YES" center is greater than "NO" center. This information is used to classify the TaqMan reactions "YES" and "NO" by comparing the fuzzy partition matrix U. If $\boldsymbol{y}_2$ represents the "YES" center and $\boldsymbol{y}_1$ represent the "NO" center (note that $\mathbf{y}_2$ is not always represent the "YES" center when the FCM algorithm is implemented), then $\boldsymbol{y}_2 > \boldsymbol{y}_1$. For example, if μ_{11} and μ_{12} equals to 0.6 and 0.4, respectively, the "YES" and "NO" reaction can be determined by following this rule:

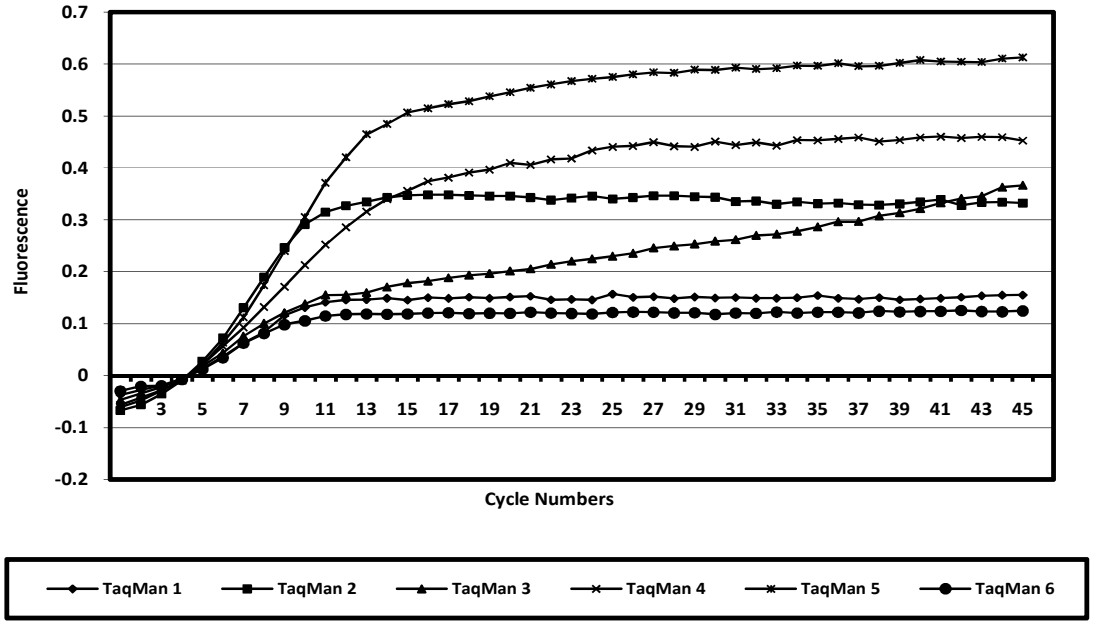

Figure 3. Output of real-time PCR. Reactions 1 to 6 indicate the $[(|V|-2)^2-(|V|-2)]/2$ TaqMan(v_0, v_k, v_l) reactions of the input instance.

if ($\boldsymbol{y}_1 > \boldsymbol{y}_2$ and $\mu_{1j} > \mu_{2j}$) or ($\boldsymbol{y}_2 > \boldsymbol{y}_1$ and $\mu_{2j} > \mu_{1j}$))
$\boldsymbol{x}_j$ ="YES"
else $\boldsymbol{x}_j$ ="NO"

Based on the proposed rule, $\boldsymbol{x}_j$ is classified as "NO" reaction since $\mu_{1j} > \mu_{2j}$ and $\boldsymbol{y}_1 < \boldsymbol{y}_2$. This rule is applied to the remaining "YES" and "NO" reactions. The whole classification process for FCM can be described in the following steps

FCM
Step 1: Initialize the membership matrix U with random values, subject to (2) and (3)
Step 2: Calculate the cluster center Y by using (4)
Step 3: Update fuzzy partition matrix U by using (5)
Step 4: Stop if $\|U(t+1)-U(t)\| < e$, otherwise go to step 2
Step 5: Determine "YES" and "NO" centers (either $\boldsymbol{y}_1 > \boldsymbol{y}_2$ or $\boldsymbol{y}_2 > \boldsymbol{y}_1$)
Step 6: Classify each TaqMan reactions by using the predefined rule

6. Results and discussion

As discussed previously, in the *in vitro* phase of the readout approach, each real-time PCR reaction is mapped to a binary output (i.e., either "YES" or "NO"), based on the occurrence or absence of an exponential amplification. Given the existence of this mapping, the subsequent *in silico* information processing is capable of determining the Hamiltonian path of the input instance (e.g., $V_0 \rightarrow V_2 \rightarrow V_4 \rightarrow V_1 \rightarrow V_3 \rightarrow V_5$, for the example instance).

FCM are implemented to classify the TaqMan reaction produced by the DNA Engine Opticon 2. The clustering parameters used are e=0.00001, m=2, N=6, and C=2. The algorithms are validated first on Matlab 7.0., which producing results shown in Figure 4. The fuzzy partition membership values for FCM clustering algorithm is listed in Table 1. Based on Table 1, it was shown that FCM algorithm can correctly classify the "YES" and "NO" reactions compared to the manual observation.

The *SILICOLIGHT* software is built based on Visual C++ 6.0. This software is able to analyze only 6 nodes of HPP problem. As shown in Figure 5, the exported data can be plot back in this software to visualize the difference between the "YES" and "NO" reaction. This software is able to cluster the TaqMan reactions, by implementing the FCM algorithm based on the results shown by using Matlab 7.0. The *in silico* algorithm to obtain the Hamiltonian Path is also implemented in this software.

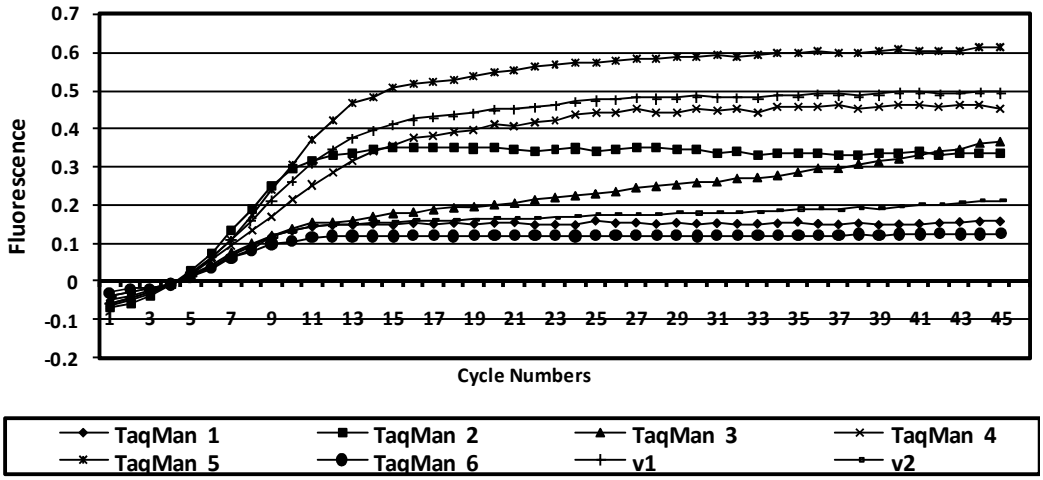

Figure 4. Output of real-time PCR. with "YES" and "NO" centers calculated using FCM clustering algorithm

Table 1. Fuzzy Partition Value for each TaqMan Reactions

TaqMan	μ_{1j}	μ_{2j}	Manual observation	Reaction (y_1>y_2)
1	0.009751	0.99025	"NO"	"NO"
2	0.63009	0.36991	"YES"	"YES"
3	0.14071	0.85929	"NO"	"NO"
4	0.97123	0.028771	"YES"	"YES"
5	0.9331	0.0669	"YES"	"YES"
6	0.027844	0.97216	"NO"	"NO"

7. Conclusion

This study offers an automation of real-time PCR-based readout approach for DNA computing, which is implemented on LightCycler System. According to the experimental results, the amplification response for "YES" and "NO" reactions can be clearly distinguished. The *in silico* information of the real-time PCR readout method is computerized in the *SILICOLIGHT* to cluster the TaqMan reaction the Hamiltonian Path automatically. By applying the FCM clustering algorithm on the output results of real-time PCR, two different TaqMan reactions, "YES" and "NO", can be clearly distinguished. Other algorithms that can handle noises and outliers can also be implemented to automatically classify the output of real-time PCR. Implementation of FCM can also be discussed further so that the implementation can be fully maximized.

8. Acknowledgement

This research is supported financially by the Ministry of Science, Technology, and Innovation (MOSTI), Malaysia under eScienceFund Research Funding (Vot 79033) and INTEL Corporation Research Funding (Vot 73332).

9. References

[1] L. Overbergh, "The use of real-time reverse transcriptase PCR for the quantification of cytokine gene expression," *Journal of Biomolecular Techniques*, vol. 14, pp. 557-559, 2003.

[2] N. J. Walker, "A technique whose time has come," *Science*, vol. 296, pp. 557-559, 2002.

[3] J. R. Lakowicz, *Principles of fluorescence spectroscopy,* 2nd ed., Kluwer Academic/Plenum Publishers, New York, 1999.

[4] Z. Ibrahim, J. A. Rose, Y. Tsuboi, O. Ono, and M. Khalid, "A New Readout Approach in DNA Computing Based on Real-Time PCR with TaqMan Probes," *Lecture Notes in Computer Science (LNCS),* vol. 4287, C. Mao and T. Yokomori, Ed. Springer-Verlag,, 2006, pp. 350-359.

[5] L. M. Adleman, "Molecular Computation of Solutions to Combinatorial Problems," *Science*, vol. 266, pp. 1021-1024, 1994.

[6] J. A. Rose, "The Effect of Uniform Melting Temperatures on the Efficiency of DNA Computing," *DIMACS Workshop on DNA Based Computers III*, pp. 35-42, 1997.

[7] D. H. Wood, C. L. T. Clelland, and C. Bancroft, "Universal biochip readout of directed Hamiltonian path problems," *Lecture Notes in Computer Science*, vol. 2568, pp. 168-181, 1999.

[8] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.