Image data analysis in qPCR: A method for smart analysis of DNA amplification

CrossMark

Massimo Orazio Spata *, Maria Eloisa Castagna, Sabrina Conoci

STMicroelectronics, Department: IPD&AMS – Research & Development, Stradale Primosole 50, 95121 Catania, Italy

ARTICLE INFO

ABSTRACT

In this paper, a method for the direct quantitative analysis of amplified DNA via q-Polymerase Chain Reaction (qPCR) in miniaturised silicon-based chip system is presented. The designed tool presented here allows the automatic extraction of meaningful information from input fluorescent images by means of digital image processing algorithm. In particular, a smart mathematical model, optimizing the integration of all the analysis steps of the fluorescence data from on chip multiple real time PCR, is described. Such a tool is able to load the digital input images, select and smartly detect the region of interest for fluorescence, elaborate the data input, calculate the average fluorescence values and finally, plot the fitted curve as output, giving also, for each well, both the Cycle Threshold (C_T) and Slope parameters.

1. Introduction

In the last years, there has been a huge interest in the development of portable systems for DNA analysis, the so-called Biochip. These devices are miniaturised systems able to integrate *on-chip* all the basic functions for biochemical analysis combining consolidated process technologies (i.e. those silicon-based) with microfluidics functions such as transport, dispensing and mixing. The main cause of such a trend is mainly related to both the possibility to be massively produced (thanks to the consolidated production technologies) and their capability to give fast and reliable results in the analysis (thanks to the precise controls of the physical parameters of the DNA analysis). [1–4]. Due to these advantages, biochips are expected to revolutionize clinical diagnosis of diseases, enabling the possibility to make fast diagnosis in the immediate closeness of patient (Point-of-Care, PoC) in a wide range of applications including oncology, drug discovery and infectious diseases [5–8].

Among the biotechnological methods used to analyse DNA, the qPCR (quantitative Polymerase Chain Reaction) is the most employed in the molecular diagnostic area. It is a biochemical technology able to generate thousands of million copies of a specific DNA segment through a thermal amplification process via the polymerase enzyme. The kinetic curve is measured in real time by recording a fluorescent signal coming from a specific dye. This signal is plotted against the number of amplification cycles on a logarithmic scale and the Cycle Threshold (C_T) is extracted to quantify the DNA [9–12].

The use of such methodology in PoC format requires the development of portable and low-cost systems. These systems typically must integrate several technological modules such us thermal, optical, chemical and data analysis modules. The data analysis module is a combination between the detection system and the algorithm of analysis. For low cost instrument, usually charge-coupled device (CCD) detectors are employed to detect the fluorescent emission of DNA targets [1]. The structure of a typical fluorescent detection system consists of an excitation source such as a laser or a white light source or a Light Emitting Diode, light delivery optics, a narrowband interference filter to select a narrow spectrum of excitation light, light collection optics and a narrowband filter to isolate the emission spectrum of fluorochrome from its excitation spectrum, image sensor, and control and processing units.

Image analysis is a fundamental process for the final performances of the system. Actually, to avoid loss of sensitivity, the algorithm of analysis has to be robust enough to process digital images extracting fluorescence data only from specific region of the PCR reactor (Region Of Interest (ROI)) at each amplification cycle.

Here, we present an optimized algorithm able to analyse in real time multiple fluorescence data coming from a silicon chip containing 6 PCR reaction wells, during the entire amplification process. This algorithm overcomes the technological drawbacks such us the movement of drops into reaction chambers, drop evaporation or air bubbles creation that can occur during the temperature increasing at each PCR cycle. Actually, during temperature ramps into the wells drops diameter and circular shape of drops could be variable. These problems has been resolved through the application of algorithmic automatic areas detection of the Region Of Interest (ROI), accelerating data computation only on ROI areas, increasing usability, eliminating human errors due to manual

* Corresponding author.
 E-mail addresses: massimo.spata@st.com (M.O. Spata), sabrina.conoci@st.com (S. Conoci).

selection of drops into wells of biochip and decreasing image camera aberration phenomena into the wells.

2. Materials and method

The Real Time PCR data have been collected by using the miniaturised platform Q3 Real Time PCR (Fig. 1). This system is based on silicon microchip able to perform Real Time PCR analysis in compact and portable way. More specifically, the platform is composed by the following components: (a) Disposable Chip; (b) Instrument. The disposable chip operate in one-step and integrates silicon temperature sensors and heaters that allow temperature control accuracy of ± 0.2 °C, with heating rate of 15 °C/s and cooling rate of 8 °C/s. It contains 6 reaction chambers made in hybrid silicon-plastic technology designed for 10 μl of master mix (Fig. 1a). The instrument thermally and optically drives the chip during the Real Time PCR process. It has been designed a device with LED light source (center wavelength 470, 530 nm) and a camera which acquires one image for every cycle step at a specific temperature of the amplification process (Fig. 1b).

The Real Time PCR tests have been carried out by using the Hepatitis B virus (HBV) commercial kit purchased from Clonit (ref. 05960467). The kit includes the hepatitis B virus (HBV) clone complete genome (consisting in HBV genome 3.2 kbps and a plasmid PBR322 vector 3.8 kbps), primers, probes and Taq-polymerase dissolved in appropriate buffer solutions. The thermal protocol consists in 10′ @ 95 °C followed by 50 cycles 15″ @ 95 °C + 60″ @ 60 °C.

The 6-well reaction silicon chip is loaded with 6–10 μl of HBV master mix and started for the PCR cycles in the Q3 platform.

The Software algorithm has been developed by Matlab code and compiled in Windows OS.

3. Results and discussions

3.1. Image detection of ROI areas

For an optimized fluorescence analysis, it is extremely important the detection of Region Of Interest (ROI) areas in the microchip. Fig. 2 reports the chip images acquisition with 1/6 ROI selected (circle dotted-line) after 50 PCR cycles either in Bayer-pattern pixels (on the left) either in RGB (Red, Green, Blue) pixels (on the right). The alignment between the chip reaction chambers and the ROI areas of the analysis is fundamental. This step can be critical since the increasing of the temperature during the cycle experiment could generate movement

of drops into the wells and, in the worst case, drop evaporation or air bubbles. During the experiment, this can produce drop movement or drop circular diameter variation. When a fixed selection of ROI areas is employed, errors in reading fluorescence are generated.

In order to solve this problem, a new flow process has been implemented. It integrates some different algorithms and creates smart decision systems which detect automatically ROI areas computing average values on those areas and calculating PCR for every cycle.

3.2. Implemented software

The developed software is based on three main core modules:

1. Fully automatic image quantification: a set of algorithms for the digital image analysis process (module 1).
2. Numerical computation: a set of algorithms for the numerical analysis of raw data coming from Real Time PCR (module 2).
3. Graphical output: a graphical user interface to show the fluorescence curves for the PCR (module 3).

More in details, the code for the digital image analysis process executes the following steps:

- Digital Image acquisition for every cycle;
- ROI smart alignment to increase precision of data acquisition and allowing an automatic analysis flow process;
- Statistical Data calculation process through the numerical analysis process;
- Plot statistical data values of Real Time PCR to analyse the presence of sigmoid.

Fig. 3 reports the algorithm flow diagrams and methods implemented for the three above described modules.

The first step of the smart ROI analysis, is the *Edge Detection Algorithm* that execute an appropriate image elaboration through a highpass filtering used to sharpen the image and detect the edges. It includes the following steps:

- reading *image luminance*
- increasing and normalize *image contrast*
- compute *image gradients* (Gaussian Kernel Radius and Gaussian Kernel width)
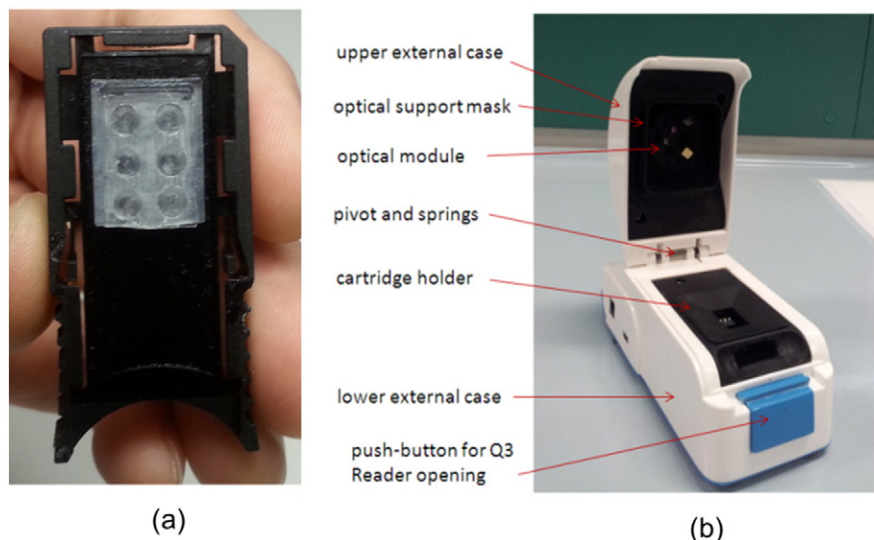

(a)

upper external case
optical support mask
optical module
pivot and springs
cartridge holder
lower external case
push-button for Q3 Reader opening

(b)

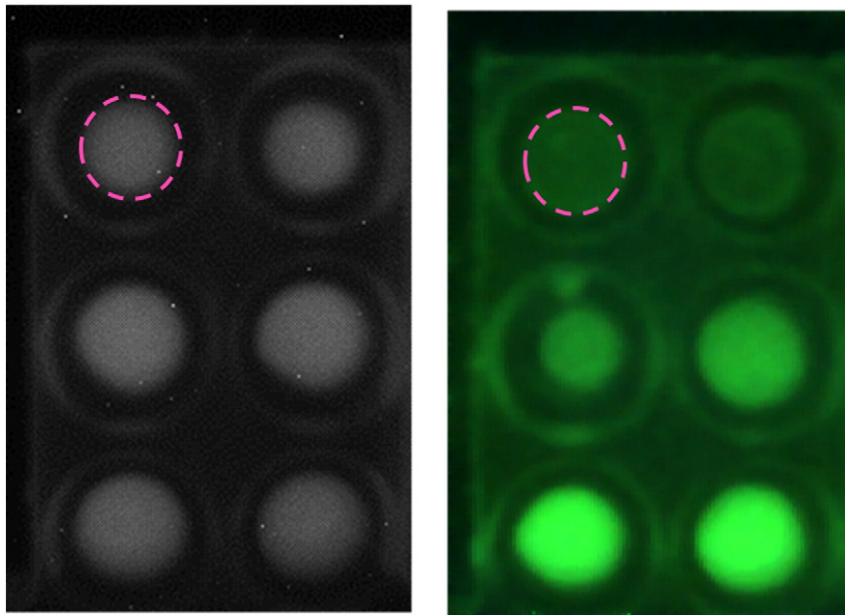**Fig. 1.** Q3 Real Time PCR Platform: (a) silicon chip; (b) Instrument.

Fig. 2. Biochip different images after 50 cycles of PCR (Bayer-pattern on the left and RGB on the right).

- perform *image hysteresis* with low and high image threshold
- write computed edges on a new image

Image luminance: is a photometric measure of the luminous intensity per unit area of light travelling in a given direction.

$$L_v = \frac{d^2\varphi_v}{dAd\Omega\cos\theta}$$

Where L is luminance, φ is luminous flux, A is area, Ω is solid angle, and θ is angle between surface normal and specified direction [16].

Image contrast: is determined by the difference in the colour (Red or Green or Blue matrix) and brightness ($(R + G + B)/3$) of the specific object in an image.

Image gradient: is a directional change in the intensity or colour in an image. Image gradients may be used to extract information from images. It is calculated through the Gaussian Kernel radius and width:

Image hysteresis: is the dependence of an image not only on its current colour level but also on its past colour level. It is based on the principle that large intensity gradients are more likely to correspond to edges than small intensity gradients. Fixing a low and high level threshold for intensity gradients, allows to find image edges.

After the *Edge Detection* computation, the application of Hough Algorithm is executed. It allows the detection of the ROI through the matchmaking of a complete specification of the exact shape of the biochip reaction well. This sub-module comprises the following steps:

- Load edge image, create on the step and use it as input image.
- Apply the Hough Algorithm using a reference image of a well with 14×14 pixel for the automatic detection and alignment.

The simplest case of Hough transform is the linear transform for detecting straight lines: $y = mx + b$ and can be graphically plotted for each pair of image points (x, y).

The equation $y = mx + b$ can be represented as a point (b, m) in the parameter space. But for computational reasons, it is therefore better to use a different pair of parameters, denoted (r, θ), for the lines in the Hough transform. These are the *Polar Coordinates*.

The parameter r represents the distance between the line and the origin, while θ is the angle of the vector from the origin to this closest point.

Using this parameterization, the equation of the line can be written as:

$$r = x\cos\theta + y\sin\theta.$$

The Hough transform algorithm uses an array, called an accumulator, to detect the existence of a line $y = mx + b$. For this line, the linear Hough transform problem has two unknown parameters: the pair (r, θ).

So for each pixel and its neighbourhood, the Hough transform algorithm determines if there is enough evidence of an edge at that pixel. By finding the highest values, typically by looking for local maxima in the accumulator space, the most likely lines can be extracted, and their approximate geometric definitions read off. Since the lines returned do not contain any length information, it is often next necessary to find which parts of the image match up with which lines.

After automatic detection of Region of Interest (ROI) described above, the average fluorescence values are calculated for all the ROI of the biochip for each PCR cycle. These average values of fluorescence are taken as input from the *numerical computation module* of the software. It has been developed via Matlab code starting from the raw data coming out from the Real Time PCR image data analysis. Basically, this code accepts as input the fluorescence data and gives as output the C_T (Cycle Threshold) values and the Slope of the PCR curves.

Fitting four-parameter sigmoidal models is one of the methods established in the analysis of quantitative real-time PCR (qPCR) data. We observed that these models are not optimal in the fitting outcome, due to the inherent constraint of symmetry around the point of inflection. Thus, we realize that it is necessary to employ a mathematical algorithm that circumvents this problem utilizing an additional parameter to accommodate the asymmetrical structures in sigmoidal qPCR data [13,15]. C_T represents the first amplification cycle in which the instrument detects a fluorescence signal due to a real amplification and it is calculated as the second derivative maximum.

The software here described has been employed for the analysis of a set of PCR data. On the basis of the literature, we have chosen a five-
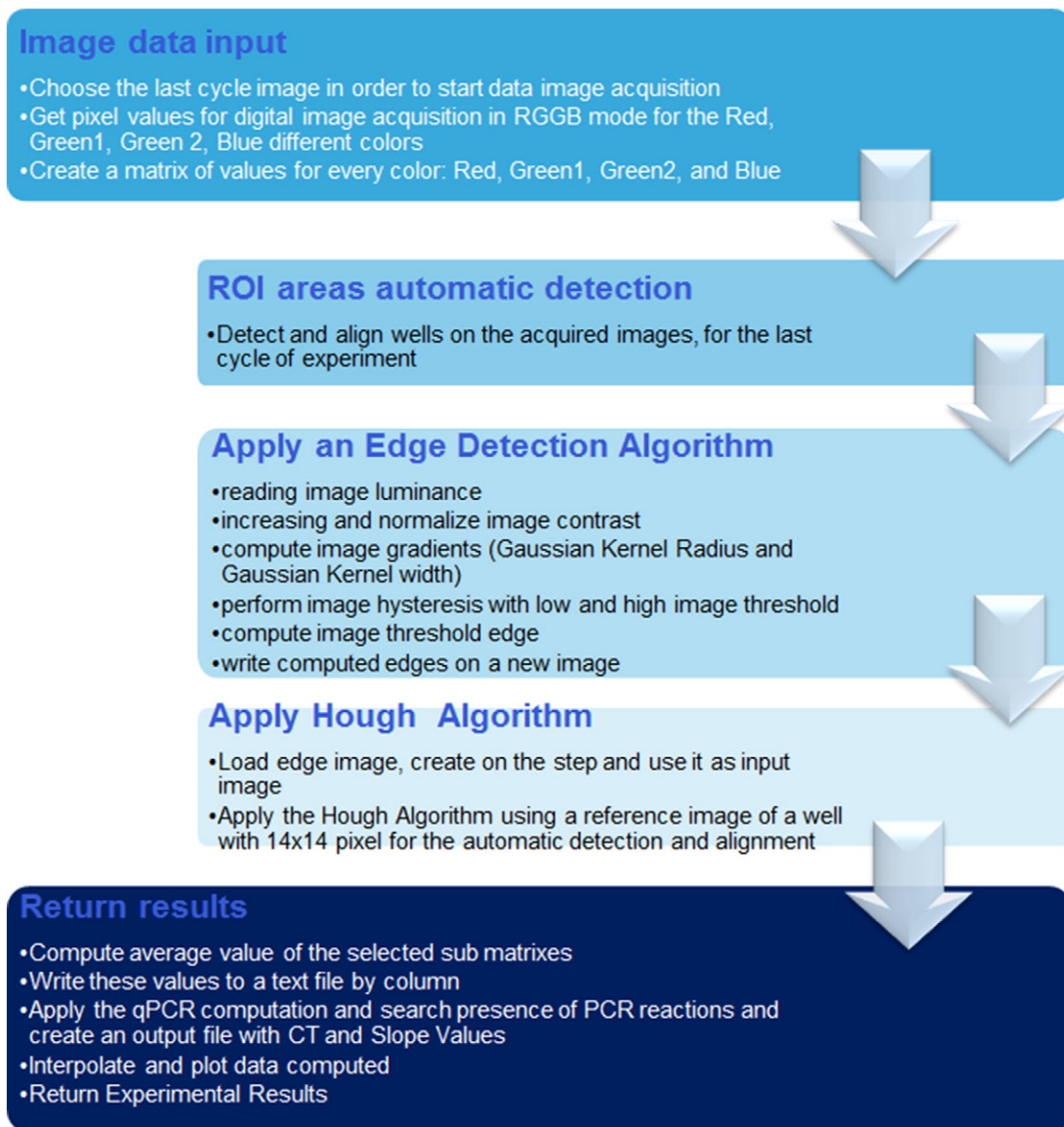
Image data input

- Choose the last cycle image in order to start data image acquisition
- Get pixel values for digital image acquisition in RGGB mode for the Red, Green1, Green 2, Blue different colors
- Create a matrix of values for every color: Red, Green1, Green2, and Blue

ROI areas automatic detection

- Detect and align wells on the acquired images, for the last cycle of experiment

Apply an Edge Detection Algorithm

- reading image luminance
- increasing and normalize image contrast
- compute image gradients (Gaussian Kernel Radius and Gaussian Kernel width)
- perform image hysteresis with low and high image threshold
- compute image threshold edge
- write computed edges on a new image

Apply Hough Algorithm

- Load edge image, create on the step and use it as input image
- Apply the Hough Algorithm using a reference image of a well with 14x14 pixel for the automatic detection and alignment

Return results

- Compute average value of the selected sub matrixes
- Write these values to a text file by column
- Apply the qPCR computation and search presence of PCR reactions and create an output file with CT and Slope Values
- Interpolate and plot data computed
- Return Experimental Results

Fig. 3. Flow diagrams of the software modules.

parameter mathematical model as fitting function [5]. The reasons of this choose has been discussed in [3], where the methods established in the analysis of quantitative real-time PCR (qPCR) data are well analysed. Therefore, the function here used for *curve fitting* of the RT-PCR raw data is the following five-parameter sigmoidal curve [13,15]:

$$g\left(x, b, c, d, k, f\right) = \mathrm{d} + \frac{c - d}{\left(1 + exp^{b(x-k)}\right)^f}$$

where:

- x is the cycle numbers
- the parameters b, c, d and k correspond to the slope, the ground asymptote, the maximum asymptote and the inflection point, respectively
- the parameter f is the additional parameter that allows to take into account the asymmetry of the lower and upper parts of the sigmoidal curve in respect to the inflection point.

By using an experimental raw dataset collected by the Q3 real time platform, we have implemented the previous formula to calculate the final values of C_T and *slope*. To do it, the following steps have been implemented:

a) Calculation of C_T

- Fitting of the raw dataset by using the previous equation, extracting the five parameters: b, c, d, k, and f. The method used to extract the 5 parameters of sigmoid curve is based on the Least Squares Method, i.e. in minimizing the norm of the difference between the output and the measured values, i.e. minimizing the residual sum-of-squares:

$$R\left(\vec{v}\right) = \frac{1}{max(1, abs(y_i))} \sum_{i=1}^{n} \left(g(x_i, \ \vec{v}) - y_i \right)^2.$$

This equation representing the residual function v which is the vector representing the 5 real parameters to be fit; $(x_i; y_i)$ are the couples of real numbers corresponding to the experimental values (cycles threshold and raw fluorescence), while $g(x_i, v_i)$ is the sigmoid function that is able to describe our output.

- Calculation of the first derivative
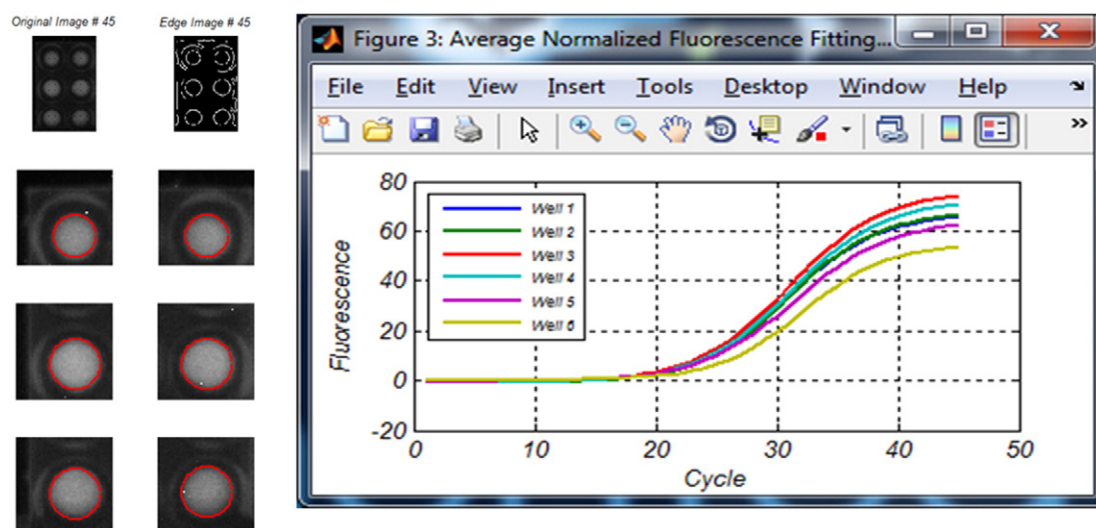- Calculation of the second derivative

Fig. 4. ROI smart image detection and Real Time PCR curves.

- Calculation of the third derivative and set this function equal to 0.0 to find roots
- Extraction of max values of the second derivative function (roots of previous function). The C_T is the *x*-value of the positive solution.

b) Calculation of the Slope

- The second derivative function has been set to 0.
- It is calculated to the inflection point.
- Evaluate the first derivative in the inflection point. This corresponds to the *slope* value.

3.3. Experimental results

Fig. 4 reports the Real Time PCR curves obtained by the software here described on the images collected by the Q3 real time platform using a silicon chip containing 6 reaction wells loaded with 6 (wells 1, 2), 8 (wells 3, 4) and 10 μl (wells 5, 6) of the HBV master mix, respectively.

It can be noticed that, the software is able to correctly analyse the data in all the PCR mix volumes employed (6 μl — wells 1, 2; 8 μl — wells 3, 4; 10 μl — wells 5, 6), even if they are not the ones optimised for the system (10 μl — see Materials and method section). Table 1 reports the correspondent C_T and Slope values obtained with the smart software here described (Smart software columns). The average C_T value is 22.4 ± 0.4 indicating that there are no relevant differences among the wells.

For comparison purposes, we also analysed the PCR curves by using the Q3 standard software (Q3 software columns). In this case, the

average C_T value was equal to 23.4 ± 0.6. This data indicates one more cycle in the C_T average value (23.4 versus 22.4) together with an increased variance (0.4 versus 0.6).

4. Conclusions

In this paper, we have described our proposed efficient image analysis algorithm using a smart automatic procedure. Our proposed algorithm use Edge detection and Hough Transform algorithm in order to optimize ROI areas data load in Biochips, and trying to decrease computational time inefficiency.

Experimental tests demonstrate that implemented software is able to calculate dynamically ROI areas and load average fluorescence values for every image cycle increasing software usability, with variable initial experiment conditions.

This algorithm overcome the technological drawbacks such us the movement of drops into reaction chambers, drop evaporation or air bubbles creation that can occur during the temperature increasing at each PCR cycle.

In the near future, this research can also be applied to robotics or to Image processing vision system [14] to automate shape recognition avoiding human evaluation errors.

Acknowledgements

Table 1
C_T and Slope values.

	Smart software		Q3 software	
	C_T	Slope	C_T	Slope
Well 1	22	1.8521	23.1	Not available
Well 2	22.4	1.8547	23.4	Not available
Well 3	21.9	2.0966	22.8	Not available
Well 4	22.2	1.9661	23.1	Not available
Well 5	22.6	1.734	23.5	Not available
Well 6	23.1	1.425	24.6	Not available
Average	22.36667		23.41667	
Standard Dev	0.402768		0.575664	

References

[1] B. Foglieni, A. Brisci, F.S. Biagio, P. Di Pietro, S. Petralia, S. Conoci, M. Ferrari, L. Cremonesi, Molecular diagnostics on a Lab-on-Chip device. A new advanced solution for mutation detection in gene-diseases, Clin. Chem. Lab. Med. 48 (2010) 329–336.

[2] M. Palmieri, E. Alessi, S. Conoci, M. Marchi, G. Panvini, Develop the "In-Check" platform for diagnostic applications, Proc. SPIE, Microfluidics, BioMEMS, and Medical Microsystems VI, Vol. 6886 2008, p. 6886-01.

[3] C. Zhang, J. Xu, W. Ma, W. Zheng, PCR microfluidic devices for DNA amplification, Biotechnol. Adv. 24 (2006) 234–284.

[4] C.J. Easley, L.A. Legendre, J.P. Landers, J.P. Ferrance, Rapid DNA amplification in glass microdevices, Methods Biol. 339 (2006) 217–232.

[5] P.-A. Auroux, D. Iossifidis, D.R. Reyes, A. Manz, Micro total analysis systems. Analytical standard operations and applications, Anal. Chem. 74 (2002) 2637–2652.

[6] R. Davidsson, F. Genin, M. Bengtsson, T. Laurell, Microfluidic biosensing systems. Part I. Development and optimisation of enzymatic chemiluminescent μ-biosensors based on silicon microchips, J. Lab. Chip. 4 (2004) 481–487.

[7] S. Petralia, R. Verardo, E. Klaric, S. Cavallaro, E. Alessi, C. Schneider, In-check system: a highly integrated silicon Lab-on-Chip for sample preparation, PCR amplification and microarray detection of nucleic acids directly from biological samples, Sensors Actuators B Chem. 187 (2013) 99–105.

[8] G. Ventimiglia, S. Petralia, Recent advances in DNA microarray technology: an overview on production strategies, detection methods and applications, Bionanoscience 3 (2013) 428–450.

[9] M.A. Valasek, J.J. Repa, The Power of Real-time PCR, vol. 29 no. 3Advances in Physiology Education Published, September 1 2005 151–159.

[10] M.J. Espy*, J.R. Uhl, L.M. Sloan, S.P. Buckwalter, M.F. Jones, E.A. Vetter, J.D.C. Yao, N.L. Wengenack, J.E. Rosenblatt, F.R. Cockerill III, T.F. Smith, Real-Time PCR in clinical microbiology: applications for routine laboratory testing, Clin. Microbiol. Rev. 19 (1) (January 2006) 165–256.

[11] W.H. Minhass, P. Pop, J. Madsen, Compilers: system-level modeling and synthesis of flow-based microfluidic biochips Architectures and Synthesis for Embedded Systems (CASES), Proceedings of the 14th International Conference on — Publication Year: 2011 2011, pp. 225–233.

[12] C. Maramis, A. Delopoulos, Efficient Quantitative Information Extraction from PCR-RFLP Gel Electrophoresis Images Pattern Recognition (ICPR), 20th International Conference on Digital Object Identifier: 10.1109/ICPR.2010.627 Publication Year: 2010 2010, pp. 2560–2563.

[13] A.-N. Spiess, C. Feig, C. Ritz, Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry, BMC Bioinforma. 9 (2008) 221.

[14] A. Elouardi, S. Bouaziz, A. Dupret, J.O. Klein, R. Reynaud, Image Processing Vision System Implementing a Smart Sensor, Instrumentation and Measurement Technology Conference, IMTC 04. Proceedings of the 21st IEEE Volume: 1 Digital Object Identifier, Vol. 12004 445–450, http://dx.doi.org/10.1109/IMTC.2004.1351085 (Publication Year: 2004).

[15] R.G. Rutledge, Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications, Nucleic Acids Res. 32 (22) (2004), e178.

[16] D. Ziou, S. Tabbone, Edge detection techniques — an overview, Int. J. Pattern Recognit. Image Anal. 8 (1998) 537–559.