Automatic Reading System for On-off Type DNA Chip

Munho Ryu*, Jong Dae Kim, and Jongwon Kim*****

Abstract: In this study we propose an automatic reading system for diagnostic DNA chips. We define a general specification for an automatic reading system and propose a possible implementation method. The proposed system performs the whole reading process automatically without any user intervention, covering image acquisition, image analysis, and report generation. We applied the system for the automatic report generation of a commercialized DNA chip for cervical cancer detection. The fluorescence image of the hybridization result was acquired with a GenePix™ scanner using its library running in HTML pages. The processing of the acquired image and the report generation were executed by a component object module programmed with Microsoft Visual C++ 6.0. To generate the report document, we made an HWP 2002 document template with marker strings that were supposed to be searched and replaced with the corresponding information such as patient information and diagnosis results. The proposed system generates the report document by reading the template and changing the marker strings with the resultant contents. The system is expected to facilitate the usage of a diagnostic DNA chip for mass screening by the automation of a conventional manual reading process, shortening its processing time, and quantifying the reading criteria.

Keywords: *DNA chip, Automatic Reading, Report Generation, Image Processing, HPV DNA chip*

1. Introduction

The diagnostic process with a DNA chip consists of a biochemical process and a reading process: the former hybridizes a DNA sample with a DNA chip and the latter scans the chip and diagnoses its clinical state [1, 2]. In detail, the reading process consists of image acquisition by scanning the chip, reading with the naked eye or image processing to detect the existence of specific genotypes, and report document generation. To accelerate DNA chip application for mass screening, the automation of the reading process is an important issue, and there have been intensive efforts to simplify the biochemical process, such as lab-on-a-chip [3].

It is necessary to develop a special reading system to automate the reading process for the following reasons. Most scanner software applications for DNA chips provide an imaging function and several analysis tools [4,5]. However, those scanner programs are usually not sufficient for the full automation of the reading process. This is because the locating and classifying of the probe spots are closely related to the specific chip design. Furthermore, as the automatically generated report should include the patient information and the chip images, some manual operations are inevitable with the scanner software provided with the scanners. To promote the usage of DNA chips, it is essential that the generated report is compatible with a prevalent commercial word processor. It will furnish the customers of the chip with a unique or flexible design of the report form without any additional modification of the automatic reading system. Any additional information besides the basic data can be easily added or configured by the user.

Although it is necessary to automate the reading process (including report generation), many studies on automatic reading have not included the whole process [6-8]. They are mainly focused on image processing for spot alignment and spot recognition. However, they do not provide the automation of the whole batch process.

In this study, we propose a system which automates all of the reading process, from image acquisition to the generation of the final report. It is implemented for a commercialized DNA chip used in cervical cancer detection - the HPVDNAChip (Biomedlab Co., Korea). The chip is scanned by GenePix (Axon Instruments, Union City, CA), which is one of the most widely used scanner programs. The proposed system automates the whole reading process, starting from the time that a DNA chip is loaded into the scanner until its report document is generated. In order to ensure compatibility with one of the most commonly used word processors, we utilized a report document template of HWP 2002 (Haansoft Inc., Korea). To generate the report, we made an HWP 2002 document template with marker strings that were supposed to be searched and replaced with the corresponding information, such as patient information and diagnosis results.

Manuscript received October 12, 2006; accepted December 19, 2006.
Corresponding Author: Munho Ryu
* Div. of Bionics and Bioinformatics, Chonbuk National University, Jeonju, Korea (mhryu@chonbuk.ac.kr)
* Bioengineering Research Center for the Aged, Chonbuk National University, Jeonju, Korea (mhryu@chonbuk.ac.kr)
** Div. of Information Engineering & Telecommunications, Hallym University, Chunchon, Korea (kimjd@hallym.ac.kr)
*** Biomedlab Co., Ltd., Ansan, Korea (jwkim@bmelab.com)

2. Requirement Specification and Design of Automatic Reading System

2.1 System Requirement Specification and Design Method

The requirement specifications of the automatic reading system are as follows:

(1) Automation without any user intervention includes the acquisition and processing of the chip image and the report generation.
(2) It provides a proper customized measure for the given diagnostic DNA chip - for locating probe spots and detecting their hybridization states.
(3) It has the input mechanism of the patient information for the report generation.
(4) The report includes the chip images, detection results on the existence of specific genotypes, and the patient information.
(5) The report form can be easily configured and edited by the users, and is compatible with an easily available word processor.

The first requirement is implemented by batching the whole process from the operation of the scanner program for image acquisition to the generation of the report. All the scanner parameters for imaging and scanning area are automatically set according to the condition and format of the targeted DNA chip. The second requirement is implemented by a separate program that reads the acquired image, locates the reference position of the probe spots, and detects these spots on the referenced positions. To fulfill this requirement, it is important to develop a measure that maximizes diagnosis performance for the targeted DNA chip. The third requirement is not directly related with the reading process itself, but it allows the user to input patient information at the proper time. Finally, the fourth and fifth requirements are implemented by generating a report from a document template based on a commonly available commercial word processor. The template should have simple rules for the users to configure the format and contents of the report easily.

2.2 HPVDNAChip for Cervical Cancer Detection

The HPVDNAChip (Biomedlab Co., Korea), designed to detect human papillomavirus (HPV) infection, one of the main causes of cervical cancer, is configured as shown in Fig. 1 [9,10]. Each chip slide has four chambers for four patients, one for each patient. Each chamber has two identical spot sets to increase diagnostic credibility. A spot set has four positive control markers and twenty-two pairs of HPV type-specific oligonucleotide probes. Each HPV type probe is also duplicated to form a pair of spots, yielding four spots for one type of probe in one chamber. The four positive control markers in each set are oligonucleotide probes of human-globin. These markers are used for test verification and locating the reference of a spot set.

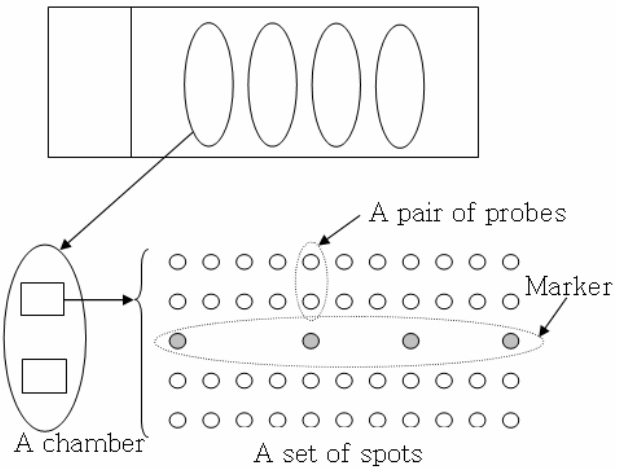

Fig. 1. The architecture of a HPVDNAChip

2.3 Diagnostic Measure

The template matching method in our previous studies has shown reasonable performance in locating marker spots for an on-off type of DNA chip [11-13]. The integration of a template matching method with prior knowledge of the shape and arrangement of the spots was successfully applied to marker-spot locating and probe-spot detection. A non-linear matching measure, called a counting measure, was proposed and proven to show better performance compared to the classical normalized covariance. The counting measure is obtained by grouping the pixels in a template region into white and black pixels and taking the difference between the numbers of white pixels in the object and background areas of the template, as shown in the following equation:

$$Counting(i,j) = \#\{(k,l) \mid I(i+k, j+l) > Th,\ (k,l) \in O\} - \#\{(k,l) \mid I(i \mid k, j+l) > Th, (k,l) \in B\} \quad (1)$$

where *I(i,j)* is the intensity of the image at the pixel position *(i,j)*. #{}, *O*, *B*, and *Th* are the number of elements of the set, the object area, the background, and threshold, respectively. The threshold is chosen to equate the white and black pixels in the template.

3. The Implementation of the Automatic Reading System

As a practical example, we implemented a system for a commercialized DNA chip for cervical cancer detection, HPVDNAChip (Biomedlab Co., Korea). The chip was scanned by GenePix™ (Axon Instruments, Union City, CA), which is one of the most widely used scanner programs. GenePix provides an excellent interface that can be implemented in HTML and scripting language. We were able to easily realize the user interface for the patient

information and triggering automatic reading process. The images were acquired through the GenePix library. We implemented a COM (component object module) for processing the acquired images and the generation of the report. The COM was programmed with Microsoft Visual C++ 6.0 and controlled by including a scripting engine in the HTML user interface program (UI/Scanner module in Fig. 2).

To generate the report document, we prepared a template in an HWP 2002 word processor, and inserted predefined marker strings corresponding to essential information. The COM application reads the template and replaces the marker strings with the appropriate contents such as the patient information or the processing results for the chip images.

Fig. 2 shows the data flow diagram. For a clearer description, each numbered flow in the figure will be inserted at the end of the steps of the following operational procedure.

AutomaticReading Procedure

1. The user loads a DNA chip into the scanner, inputs the patient information, and pushes the start button on the user interface page in the UI/Scanner module (②).
2. The UI/Scanner module scans predefined areas that are large enough to include eight spot sets for four chambers, and saves them as image files. The image files are saved in the predefined folder, and have unique names composed of the patient's name and a unique key from the patient information. The key number is selected by the chip users from uniquely identifiable numbers such as the registered number (①, ②).
3. The UI/Scanner module calls the COM server to trigger the reading process. At this time, the names of the image files are transferred to the COM server (③).
4. The COM server reads the report document template to generate (④).

Do For Four Chambers.

5. The COM server reads two images of the spot sets for a chamber, locates the marker spots, and detects the probe spots (④). The basic idea for locating marker and probe detection is the template matching method, which is specified in the references [11-13].
6. Patient information, the probe spot detection results, and the two images of the probe sets of optimized size are inserted into the report document template by the replacement method described earlier. The resultant report is saved in the predefined directory with the unique filename of step 2 (⑤).

End Do

7 The COM notifies the UI/Scanner module of the success remarks for all four chambers to be displayed (⑥).
8. The UI/Scanner module displays the remarks on the comment area in the UI page.

End AutomaticReading

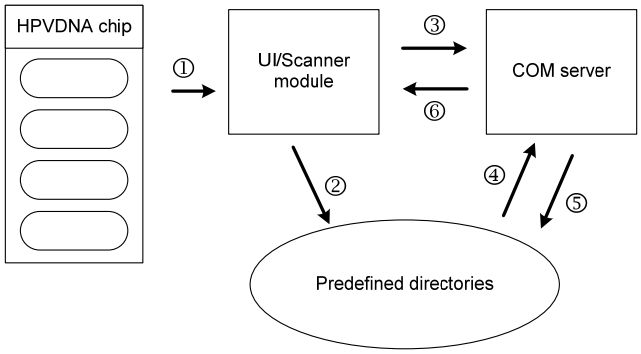

Fig.2. Data flow diagram

Fig. 3 shows a portion of the user interface page in the UI/Scanner module which triggers the automatic reading process. In this example, reading from the left-hand column, the diagnosis register number, patient information (name, id, sex, and age), and other information are entered in order. The four rows correspond to the four chambers in a DNA chip, respectively. After all the necessary information has been entered, the "Generate Report" button is pressed to trigger the automatic reading process as described earlier.

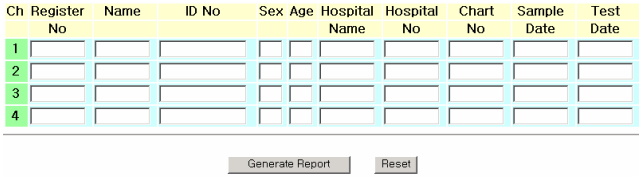

Fig.3. A portion of the GenePix page which triggers the automatic reading process

In order to give a clearer presentation for report generation, the binary views of the report document template and the generated report are shown in Fig.s 4 and 5. They are both in HWP-2002 formats and opened in the Binary Editor Window of the Microsoft Visual Studio, which provides three vertical regions of views: the leftmost shows the start address, and the center and rightmost regions show the hexadecimal and the ASCII format views of the contents, respectively. The marker string corresponding to genotype number 33 was inserted as '!33', as in the highlighted area in Fig. 4. In this example, the patient was infected with the type 33 of HPV and the corresponding probe was detected as positive. The marker string was replaced with a check mark as in Fig. 5. The COM server reads the report document template and searches and replaces the marker string to be replaced with the analyzed information content.

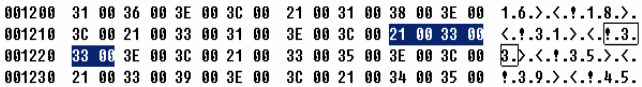

Fig. 4. Binary view of the report document template that shows the predefined marker string '!33'

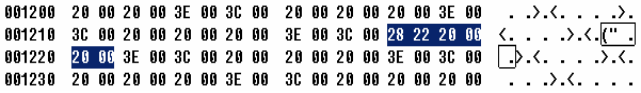

Fig.5. Binary view of the generated report that contains the check mark instead of the predefined marker string '!33'. In this example, the patient was infected by the type 33 of HPV.

Fig. 6 and 7 show how the blank bitmap image is prepared in the report document template and how it is replaced with the actual image of the probe set. The blank bitmap image is inserted into the report document template. The marker string 'BMP_MARKER_1' is written at the start of the image data area as shown in Fig. 6. When the report is generated, the actual chip image replaces the original blank image starting from the searched offset as shown in Fig. 7. The other image is also processed in a similar fashion with the marker string 'BMP_MARKER_2'.

Fig. 6 Binary view of the report document template that shows the predefined marker string 'BMP_MARKER_1'

Fig. 7. Binary view of the generated report showing that the predefined marker string 'BMP_MARKER_1' was replaced with the actual chip image

Fig. 8 and 9 show the portions of the report document template and the generated report, respectively. In Fig. 8, the spot configuration of the HPVDNAChip is displayed just for a reference on the left-hand side, and two bitmap images, which are to be replaced with actual images, are displayed on the right-hand side. The reference configuration shows the spot positions of each genotype number relative to the four marker spots. On the bottom, a table is displayed to show the existence of genotypes. For example, the marker string '!33' corresponding to genotype number 33 was inserted into the report document template. Those marker strings in the starting regions in the bitmap image and genotype result text are searched and replaced with actual chip images and genotype results when a report document is generated. In Fig. 9, we can see that the two bitmap images have been replaced with the actual chip images. Note that the existence of the genotype number 33 is confirmed from the actual images, and the corresponding cell in the table on the bottom is also checked.

4. Results and Discussions

This study proposed an automatic reading system for a diagnostic DNA chip for application to a commercialized DNA chip for cervical cancer detection. The proposed system automatically fulfills the whole reading process, which comprises image acquisition, image analysis, and report document generation, without any user intervention.

There were two main reasons for which marker locating failed and the remaining process was aborted. First, the sample showed a negative control. The sample did not hybridize even with the human-globin positive control marker spot, which was used to ensure that there was no contamination around the laboratory environment. Second, the acquired image was in too poor condition to locate the marker spots. In any case, we generated a script file instead of the report, with which the chip user could locate the marker spots manually. After the manual location of the markers, the script started the remaining process automatically. As the above failures in marker locating are closely related to the DNA chip production quality or to the degree of experience of the user, they are recorded and fed back to the chip manufacturer.

The proposed system is expected to facilitate the usage of diagnostic DNA chips for mass screening through the automation of the conventional manual reading process, shortening its processing time and quantifying the reading criteria. Even though the proposed system was implemented for a specific DNA chip and a specific scanner program as an example, it can be applied to the automatic reading system for general DNA chips.

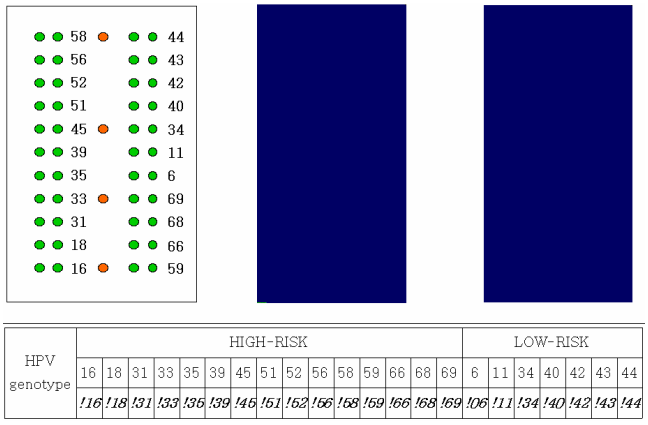

Fig.8. A portion of the report document template

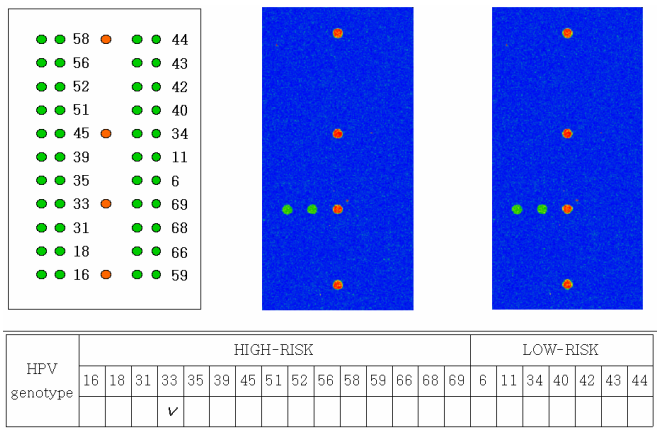

Fig.9. A portion of the report actually generated

Acknowledgments

This research paper was supported by Chonbuk National University in 2005, and by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (the Center for Healthcare Technology Development, Chonbuk National University, Jeonju 561-756, Republic of Korea).

References

[1] R.S.H. Istepanian, "Micro-array Image Processing: Current Status and Future Directions", IEEE Trans. Nanobioscience, vol.2, no.4, pp.173~175, Dec., 2003.
[2] Y. Zhou, J. Liu, "AVA: Visual Analysis of Gene Expression Micro-array Data", Bioinformatics, vol.19, no.2, pp.293~294, 2003.
[3] L.J. Ridley, "Guide to the Radiology Report", Australasian Radiology, vol. 46, pp.366-369, Dec., 2002.
[4] GenePix 4000A User's Guide, Axon Instruments, Inc., 1999.
[5] Operators' Manual, GSI Lumonics, 1999.
[6] M.R. Fielden, R. G. Halgren, E. Dere, T.R. Zacharewski, "GP3: GenePix Post-processing Program for Automated Analysis of Raw Micro-array Data", Bioinformatics, vol.18, no.5, pp.771~773, May, 2002.
[7] D. Bozinov, "Autonomous System for Web-based Micro-array Image Analysis", IEEE Transactions on Nanobioscience, vol. 2, no.4, pp.215~220, 2003.
[8] Z.Z Zhou, J.A. Stein, Q.Z Ji, "GLEAMS: A Novel Approach to High Throughput Genetic Micro-array Image Capture and Analysis," Proceedings of the International Society for Optical Engineering (SPIE), vol.4266, pp.13~23, 2001.
[9] C.J. Kim, J.K. Jeong, M. Park, T.S. Park, T.C. Park, S.E. Namkoong, J.S. Park, "HPV Oligonucleotide Microarray-based Detection of HPV Genotypes in Cervical Neoplastic Lesions," Gynecologic Oncology, vol.89, pp.210~217, 2003.
[10] B.S. Choi, O. Kim, M.S. Park, K.S. Kim, J.K. Jeong, J.S. Lee, "Genital Human Papillomavirus Genotyping by HPV Oligonucleotide Micro-array in Korean Commercial Sex Workers", Journal of Medical Virology, vol.71, pp.440~445, 2003.
[11] J.D. Kim, S.K. Kim, J.S. Cho, J. Kim, "Knowledge-based Image Processing for On-off Type DNA Micro-array", Proceedings of the International Society for Optical Engineering (SPIE), vol.4623, pp.38~46, 2002.
[12] M. Ryu, J.D. Kim, B.G. Min, "Robust Template-matching Measurements for Variations in Signal Amplitude", Journal of Optical Engineering, vol.43, no.2, pp.482~488, Feb., 2004.
[13] M. Ryu, J.D. Kim, B.G. Min, M.G. Pang, J. Kim, "Non-linear Matching Measure for the Analysis of On-off Type DNA Micro-array Images", Journal of Biomedical Optics, vol.9, no.3, pp.482~438, May, 2004.

Munho Ryu
Ryu received a BS degree in Control and Instrumentation Engineering and a PhD degree in Biomedical Engineering from Seoul National University, Korea, in 1990 and 2004, respectively. He worked for Daewoo's Heavy Industry Machine Tool Division from 1990 to 2000. He also worked for Biomedlab Co., Ltd. from 2000 to 2005. He has been a professor at Chonbuk National University since 2005. His research interests are bio-image processing and remote rehabilitation.

Jong-Dae Kim
He received a BS degree in Electronics Engineering from Seoul National University in 1982, and MS and PhD degrees in Electrical Engineering from the Korea Advanced Institute of Science and Technology, Seoul, Korea, in 1984 and 1990, respectively. He worked for Samsung Electronics from 1988 to 2000 as an electrical engineer. In 2000 he joined the Division of Information and Communication Engineering, Hallym University, as an assistant professor. His current interests are focused on image/video signal processing and bioi-maging.

Jongwon Kim
He received BS and MS degrees in Physics from Seoul National University in 1985 and 1987, respectively. He received a PhD degree in Biomedical Engineering from Seoul National University in 1992. He has been the CEO of Biomedlab Co., Ltd. since 1994. His research interests include artificial hearts and DNA chips.