Optica Applicata, Vol. XXXVIII, No. 2, 2008

Evaluation of fluorescence images acquired from oligonucleotide libraries

ANNA CYSEWSKA-SOBUSIAK*, MICHAŁ BOŁTRUKIEWICZ, JOANNA PARZYCH

Division of Metrology and Optoelectronics, Institute of Electrical Engineering and Electronics, Poznań University of Technology, Piotrowo 3a, 60-965 Poznań, Poland

*Corresponding author: Anna.Cysewska@put.poznan.pl

By using a given DNA sequence, labeled with an organic dye, fluorescence may be induced as the effect of the hybridization of the sequence with complementary oligonucleotides located in a so-called library. The analysis consists in the detection of lighting points, in which the hybridization occurred. Each image of the created matrix corresponds to the optical response of a given oligonucleotide included in the investigated sequence spectrum. A novel approach to modeling of images obtained from a virtual oligonucleotide library and detected by a CCD camera is presented. Tests to determine optimum algorithms of the oligonucleotide library detection were carried out in the form of a computer simulation. Analyzed recorded images do not represent the actual illuminated oligonucleotide library, but were obtained by simulation in an identical way and by means of the same set of instruments. As a result, it was proven how to obtain correct detection depending on a mutual location of a given oligonucleotide library and the CCD matrix.

Keywords: CCD matrix, DNA, fluorescence, image analysis.

1. Introduction

The Human Genome Project [1, 2] has generated a great increase in genetic sequence information: the development of GeneChip TM technology (Affymetrix, Santa Clara, CA) and the HP G2500A GeneArray scanner (Hewlett–Packard, Palo Alto, CA) provides the capability for rapid analysis of complex genetic information. The Human Genome Project was completed in April 2003 and the finished human genome was published concurrent with the 50th anniversary of the discovery of the double helix [3–5]. DNA is a double-stranded molecule held together by hydrogen bonds between base pairs of nucleotides. The molecule forms a double helix in which two strands of DNA spiral around one another. The molecular complementarity of the two strands of the DNA double helix makes it possible to model changes in DNA sequences manifested by related changes in their melting temperature, thermodynamic data

and optical properties [1, 2, 6–10]. Computational aspects are very important in gene analysis and sequence mapping. The scale of even the smallest genomes of biologic interest is too large to be effectively analyzed using the traditional tools of molecular biology and genetics. Combinatory tools currently available can be even better to acquire information about sequences than the tools in use to design the physical maps. However, optical and optoelectronic methods of measurements are of great and growing importance in techniques used in biology and medicine [2, 6, 11, 12]. In this paper, problems connected with modeling and simulation of optical signals acquired as fluorescence images from an array of integrated oligonucleotides are considered.

DNA is made up of subunits called nucleotides (monomers). A gene is a linearly ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (*i.e.*, a protein or RNA molecule). Each nucleotide consists of a phosphate group, a sugar, one of the bases: adenine A, thymine T, cytosine C, guanine G (Fig. 1**a**), and humans have approximately 3 billion bases in the DNA code. Any given person has approximately 9 million kilometers of DNA in their body. Molecular complementarity (A only with T and C only with G) (Fig. 1**b**) can be used in sequencing by hybridization (SBH) employed to analyze DNA. The genetic code in DNA occurs in triplets, such as ATG. The base sequence of the triplet form in the partner strand is therefore TAC. The C–G base pair is more stable than the A–T base pair; if DNA is rich in the C–G content, the greater is the energy needed to separate the strands. The optical density of a mixture of free nucleotides would be higher than the optical density of DNA of the same composition. Thus, any disruption of the duplex has to cause an increase in the optical density connected with hyperchromicity. Higher stability at the same temperature means lower optical density. Under the same conditions, the C–G pairs yield lower components of total optical density than the A–T pairs.

The nearest-neighbor model allows the prediction of optical effects of changes in the DNA duplex status [8, 9]. Sensing changes in a given fluorescently labeled N-mer is performed using a "light-directed" oligonucleotide library. Such a library consists of 4^L short testers (L-mers that are the oligonucleotides of length L) which are all known structurally and located in known positions. Only testers that are capable of full hybridization and give the brightest images, should be taken as the oligonucleotides searched for in the library. Optical signals, as consequences of thermodynamic properties of gene molecules allow the use of fluorescence techniques, easy to automate analysis [1, 2, 7–10, 12, 13]. In accordance to Beer's law, the total fluorescence intensity is proportional to the light absorbed and to fluorescence efficiency. The idea of measurements is based on the effects of the fluorescence process in which a molecule absorbs light of a given wavelength λ_1, is then excited to a higher molecular state, and decays from that state in such a way that light of a longer wavelength λ_2 is emitted [14]. Nucleotides can strongly absorb ultraviolet and blue radiation. Light in

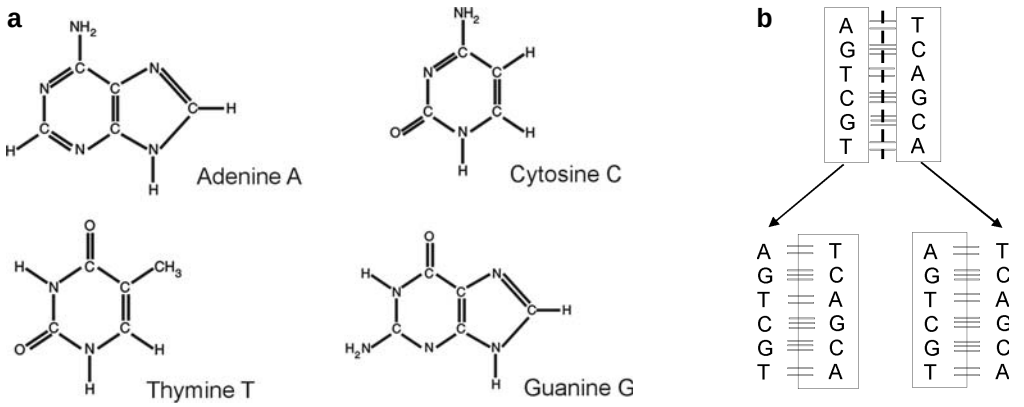

Fig. 1. Four bases which are four letters in the DNA alphabet (**a**), example of basepairing: A = T (T = A), C ≡ G (G ≡ C) (**b**).

the red, orange and yellow regions may be emitted by fluorescence and can then be detected by a sensitive photodetector.

After the fluorescently labeled DNA sequence has been banded onto a set of testers located in the integrated library, the fluorescence peaks may be detected by analysis of the obtained image characteristics. The subject of this paper is focused on problems connected with the evaluation of parameters specifying the response of a virtual, integrated oligonucleotide library. The presented analysis is concerned with images detected by a CCD camera. Reasons for the interest in using CCD include qualitative as well as quantitative analysis, ease of operation, real-time operation, along with sensitivity and precision of detection. However, the fundamental limitation of CCD detection [15–17] remains the image smear which needs to be eradicated and the process improved. Based on the knowledge of elementary luminous technology phenomena, it is noticeable that particular active library cells not only light CCD matrix converters located directly over the hybridizing element, but also (though to a lesser extent) light the adjoining converters. The images acquired in this way are highly blurred and difficult to be explicitly decoded.

2. Mutual location of the CCD matrix and the oligonucleotide library

2.1. Photometric dependences

The tests to determine optimum algorithms of oligonucleotide library detection were carried out in the form of a computer simulation. The following assumptions related to the structure of the measuring point and experiment conditions were taken:

– The size of library cells is identical to that of the light-sensitive elements of the CCD camera;

– The simulated camera area contains 10×10 CCD converters;

– Each converter quantizes the measured light intensity on 16 levels;

– The camera and library are situated such that the CCD converters are located over the corresponding library cells;

– All oligonucleotides creating a single cell fully hybridize, which permits the assumption of the identical luminance of the entire cell surface;

– The library temperature allows all oligonucleotides to hybridize;

– All the hybridizing oligonucleotides, disregarding their structures emit radiation of identical intensity;

– The CCD camera and oligonucleotide library are in the same medium, therefore no light refraction occurs.

Particular CCD converters register intensities of light received from library cells. Determination of these intensities is based on fundamental rules of photometry [18]. The oligonucleotide library analysis consists in detection of points, in which hybridization occurred. The Figure 2**a** illustrates illumination of the CCD matrix by a single lighting site in the 100-cell oligonucleotide library, while Fig. 2**b** shows the image recorded by a CCD matrix containing 10×10 converters. Interfering illumination from surrounding lighting sites may fog the image and cause a false interpretation. The phenomenon observed may be considered as blur.

The illumination intensity recorded by the particular converters was determined by known photometric rules. Assuming small area A of the cell, the luminous flux Φ penetrating a hypothetical plane A' (Fig. 2**a**), of an area equal to that of the library cell, deviating from normal to its surface by angle α is then:

$$\Phi = IA\cos\alpha \tag{1}$$

of which the radiation intensity I emitted by the hybridizing element has been assumed as known. Then, the light intensity recorded by a CCD converter, actuated by flux inclined at angle α to normal to the CCD area, will be:

$$E = \frac{\Phi}{P}\cos\alpha = \frac{IA\cos\alpha}{P}\cos\alpha = I\cos^2\alpha \tag{2}$$

of which E is the radiation intensity recorded by the CCD matrix, P is the area of the light-sensitive element of the CCD converter (equal in size to that of the element $P = A$). Because of the geometric dependences resulting from Fig. 2**a** and known trigonometrical relationships, dependence (2) can be simplified to:

$$E = I\frac{d^2}{d^2 + x^2 + y^2} \tag{3}$$

The effect on the adjacent converters may be significantly limited by reduction of the distance d between an oligonucleotide library and the CCD matrix. This process leads to increased sharpness of the image. Due to the fact that the possibility of changes in the camera optical system parameters has been assumed, the sharpness adjustment

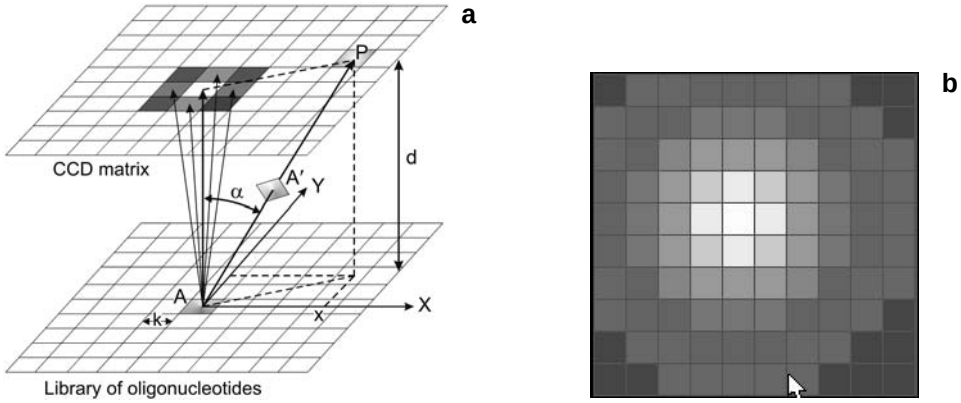

Fig. 2. Exciting a CCD matrix by a single lighting site in the 100-cell oligonucleotide library (**a**), image registered by a CCD matrix containing 10×10 converters, at a single lighting cell of the library (**b**).

process does not require any physical change of distance, simply a selection of the proper focal length for the lens. During the tests, the ”sharpness adjustment” was simulated by changing the distance *d*. Due to the proven efficiency of the method used, image sharpness adjustment by means of digital filtration was preliminarily withdrawn.

If the CCD matrix surface illuminated by a particular element significantly exceeds its size, a characteristic luminous spot is created and a clearly marked maximum is observed in the signal detection area.

2.2. Stages of sensing changes in library optical response

The fact that DNA can be denatured and renatured is a basis for predicting the optical effects of hybridization, depending on base pairing between independent complementary

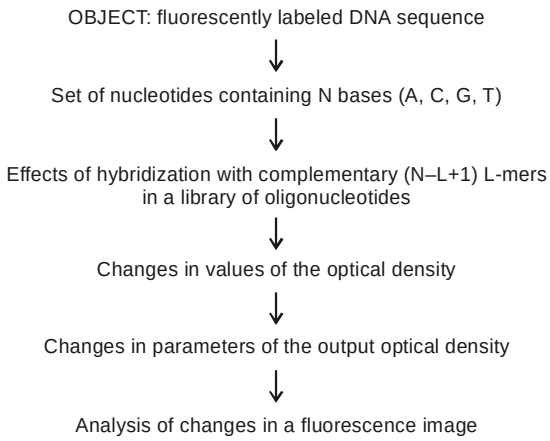

Fig. 3. Relationships specifying fluorescence sensing changes in a DNA sequence structure.

single strands. Relationships specifying fluorescence sensing changes in the DNA sequence structure are shown in Fig. 3.

There are a number of current chemical compounds which may be used as fluorescent labels because of appropriate spectral bandwidth. All the organic dyes have specific absorption peaks at specific wavelengths. Selection of a suitable dye is very important from the point of view of the sensitivity and resolution. In accordance with Stoke's law, the difference between excitation and emission wavelength should be as large as possible [14]. On the one hand, the fluorescence signal intensity is proportional to the amount of dye, and on the other hand it very much depends on the sequence order, as well as the extent of hybridization – when the degree of complementarity and number of G–C pairs increases, the emitted signal also increases. The fluorescence weak light must be collected to sense very low differences between levels of light intensity.

Affecting factors can interfere at different stages of processing and to different degrees. As a consequence, a loss of ability to detect changes in optical responses can appear or false end-indications can occur, giving positive or negative errors [19]. In the real sequence, the ideal number of all L-mers is equal to $N-L+1$, while in any other case when the sequence composition is changed, this number can be underestimated or overestimated. It is interpreted as negative or positive errors occurring, respectively. One of the real situations can also be induced by the above mentioned interfering illumination of a given dark site by other lighting sites. Lighting sites creating a detected image illuminate the related sensors of a CCD matrix, however, some adjacent sensors are also partly illuminated. Because the detected images will be affected by noises, there are problems with reliable decoding and interpretation.

3. Results of simulation

3.1. Specificity of changes in emitted images

In the process of fluorescence excitation, a molecule absorbs a photon of light, is then excited to a higher molecular state, and decays from that state in such a way that fluorescent light of a longer wavelength is emitted. The maximum fluorescence intensity of a given sequence of nucleotides depends very much on length and base content. A final measure for a given pixel, as the elementary image component, is the level of emitted light intensity. Important affecting factors can be, *e.g.* an image smear or mutual location of a given photodetector and examined oligonucleotide library. Generally, the A = T pairs are less stable than the C ≡ G pairs, however, the AA/TT stability equals the TT/AA stability > the AT/TA stability > TA/AT stability, and the CG/GC stability > the GC/CG stability equal to the GG/CC stability and the CC/GG stability [8]. The maximum fluorescence intensity of a given sequence of nucleotides depends on length and base content, *e.g.* the "strong" pairs C–G (equal to G–C) will shine more than the "weak" pairs A–T (equal to T–A) [12].

3.2. Simulation studies

The number of lighting sites and particular signal intensities depend on how many testing sequences there are and to what extent they hybridize with complementary fragments of a studied sequence. A given oligonucleotide will be included in the *N*-mer spectrum when the brightest variant of all *L* possible variants of brightness occurs. For the tester length equal to, *e.g.*, 6 oligonucleotides, we have $4^6 = 4096$ sequences which are positioned in the library cells (Fig. 4**a**). For example: at the sequence of length $N = 27$, the ideal number of *L*-mers is equal to $N-L+1 = 22$. Each composition of bases included in a given tester can give the light signal. If any tester, *e.g.* 2 in a given cell may fully hybridize, only this one will be taken as the detected oligonucleotide, because the obtained image is brightest when compared to all other images. If the CCD matrix surface illuminated by a particular element significantly exceeds its size, a characteristic luminous spot is created and a clearly marked maximum is observed in the signal detection area. If the CCD camera area designated to record the signal emitted by a defined element is shifted in relation to the surface it occupies in the library, then, no clearly marked maximum occurs in the area designated for signal detection, "distributed" onto a few adjacent areas. Figures 4**b** and 4**c** illustrate differences between the ideal and disturbed images, respectively.

If image sharpness is high, specific step changes will appear in light intensity caused by the pixels of extreme grey levels. With image binarization, all pixels

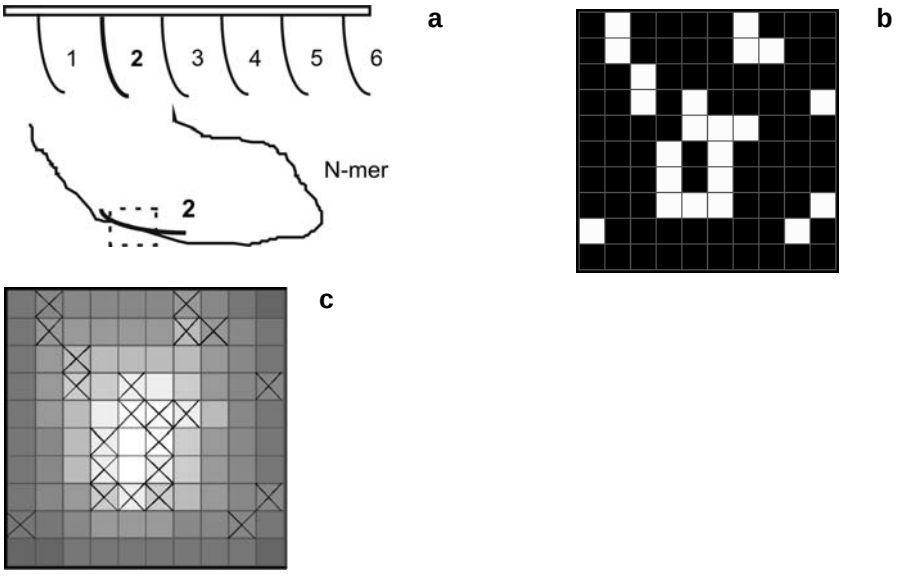

Fig. 4. Illustration of hybridization process for one of the *L*-mers included in a library with a complementary site of a given *N*-mer (**a**); example of an ideal optical response (**b**); a disturbed image obtained by means of a CCD camera: sites corresponding to effects of hybridization are marked "×", black square means a lack of signal, white square means full hybridization (**c**).

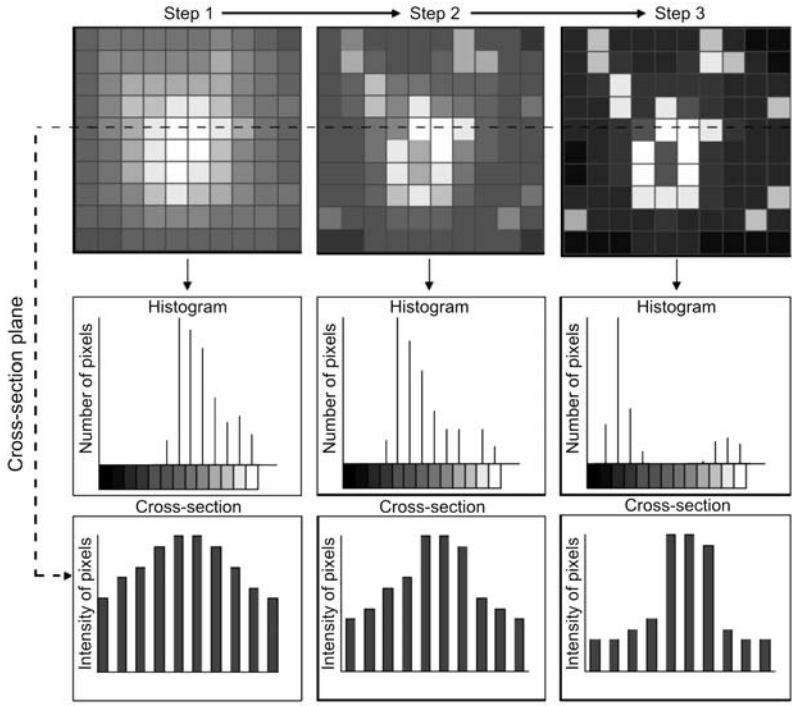

Fig. 5. Subsequent sharpness adjustment phases and image functions capable of representing valuation criteria.

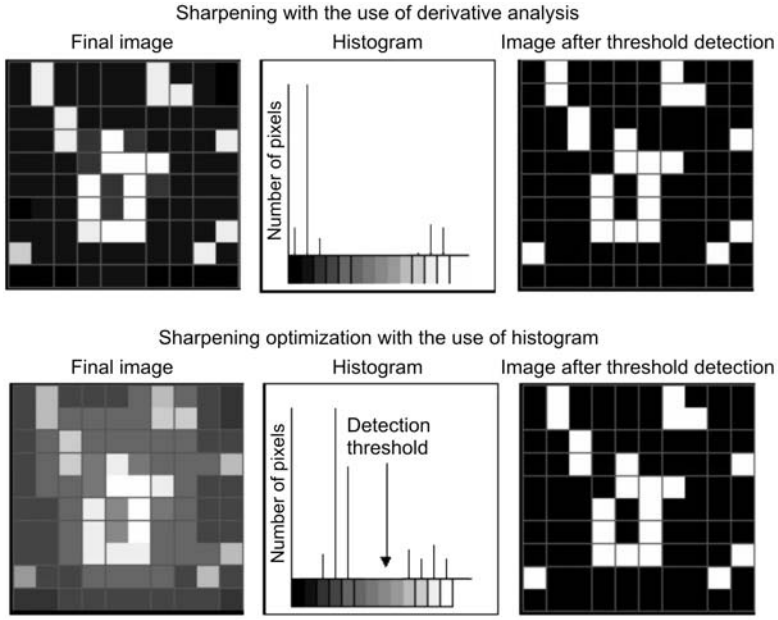

Fig. 6. Subsequent phases of sharpening for two quality assessment criteria.

may be divided into two groups: each of them contains only "illuminated" or "non-illuminated" sites, respectively. One important requirement is the determination of a proper threshold value of light intensity. Intensity of the smallest group of pixels (the histogram minimum), becomes a very effective threshold to form binarization. Important limitations are connected with many kinds of affecting factors including:

– firstly, the DNA set is very large;

– secondly, many sequences are similar in appearance of the same *L*-mer more than once;

– thirdly, a resolution of detecting changes at a level of light intensity must be high enough to distinguish particulate composition of bases.

The "recorded" images, being the input data, did not represent the actual illuminated oligonucleotide library, but were obtained by simulation in the identical manner and with the same set of instruments. The results are shown in Figs. 5–7. Figure 5 presents 3 subsequent phases of increasing sharpness. Figure 6 shows subsequent image sharpness adjustment phases carried out with the use of two criteria of image sharpening, assuming the correct mutual location of the CCD camera and the library. However, Fig. 7 concerns the influence of the mutual location of the camera and the library on the efficiency of luminous oligonucleotide detection.

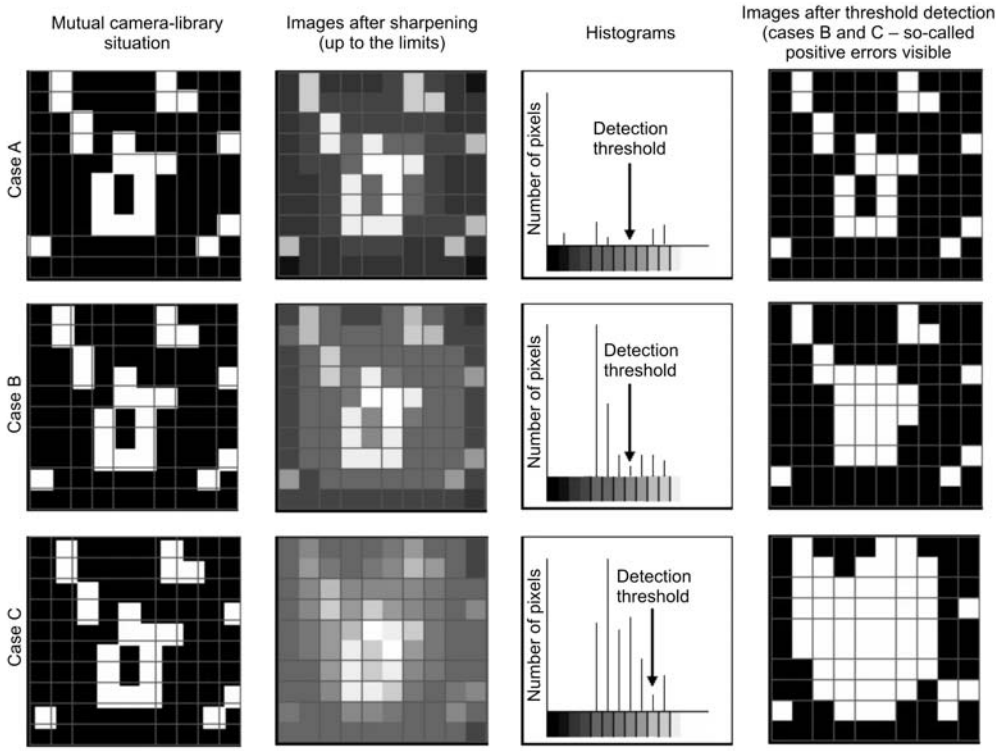

Fig. 7. Influence of the mutual location of the camera and the library on the efficiency of luminous oligonucleotide detection.

In Figure 5, diagrams of changes in the image intensity evaluated for the marked fifth horizontal line from the top and histograms are shown. Equally interesting changes are observed in the image histogram, in which the definite minimum occurs, dividing it into two distinct parts. Thus the well sharpened image is characterized by step intensity changes around the edges, caused by extreme level grey pixels present in the image (an ideal case shown in Fig. 4**b** represents two of the most extreme levels). This proves the usefulness of typical sharpness adjustment algorithms in the process. Very interesting behavior of the image histogram was observed at subsequent phases of sharpness adjustment. It allowed slight modification and concurrently optimization of the classic sharpness adjustment algorithm.

Figure 6 shows subsequent image sharpness adjustment phases carried out using efficient criteria, *i.e*. image sharpening using derivative analysis and optimization with the histogram. The case concerns the ideal oligonucleotide – CCD matrix mutual location. In reality, however, these two elements may be mutually shifted. As seen in the picture shown in Fig. 7, detection of the oligonucleotide library realized by means of sharpness adjustment in the case of significant shifts (cases B and C) appears that it is too weak.

No further effective algorithm sharpening by means of digital filtration could be found. Thus it can be said that too large mutual shifts impede effective analysis. It was proven that almost 100% effectiveness of detection is obtained, though only in the case of shifts not exceeding 15% of a CCD element size. Along with approaching the limit of the said 15% of CCD element size, grows the probability of occurrence of so-called positive errors, consisting in detection of non-active nucleotides. Positive errors consist in finding hybridization in a position in which it does not occur. Thus, the number of the fully detected hybridizing oligonucleotides is overestimated. Increasing displacements between both the investigated light-directed library and the CCD sensing array can cause the appearance of a number of the positive errors so much so that reliable acquisition of the library optical response becomes impossible.

4. Conclusions

In the paper, problems connected with the evaluation of parameters specifying the response of a virtual integrated oligonucleotide library are considered. The analysis concerns fluorescence signals to be obtained as a result of hybridization and detected by a CCD camera equipped with an adjustable focal length optical system. The considerations presented here prove the high effectiveness of the typical sharpness adjustment algorithms as basic instruments in the process of hybridizing oligonucleotide library detection. This process may also be accelerated through decimal-to-binary conversion of the image upon detection of the characteristic histogram minimum. The level and change detectability precision in the optical response emitted by the library largely depend on the relation between the measuring

fluorescence signal and background signal originating from the inducing radiation. The luminous background of the image recorded by the CCD camera may in consequence lead to positive errors consisting in finding hybridization in a position in which it does not occur. As a result of simulations carried out on oligonucleotide library – CCD matrix mutual location, it was also proven that too large shifts imply the possibility of single positive errors occurrence, the number of which grow along with the growth of the shift value up to the complete impeding of detection. In the case of image broadening with large mutual shifts of the CCD matrix and the oligonucleotide library, negative errors may also occur – the "broadened" image is interpreted as a background signal. Work in progress concern the measures that could become the useful criteria of this adjustment reliability.

References

[1] International Human Genome Sequencing Consortium, *Initial sequencing and analysis of the human genome*, Nature **409**(6822), 2001, pp. 860–921.

[2] KAISER R., *Tools for Genome Analysis, in the Biomedical Engineering Handbook*, [Ed.] J.D. Bronzino, CRC Press, Boca Raton, 1995, pp. 1489–501.

[3] WATSON J.D., CRICK F.H.C., *Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid*, Nature **171**(4356), 1953, pp. 737–8.

[4] *Publication of DNA Structure: Double Helix at the 50th Anniversary*, Nature **422**(6934), 2003.

[5] Special Section: *Building on the DNA revolution*, Science **300**(5617), 2003.

[6] LE BRIGAND K., RUSSELL R., MOREILHON C., ROUILLARD J.-M., JOST B., AMIOT F., MAGNONE V., BOLE-FEYSOT C., ROSTAGNO P., VIROLLE V., DEFAMIE V., DESSEN P., WILLIAMS G., LYONS P., RIOS G., MARI B., GULARI E., KASTNER P., GIDROL X., FREEMAN T.C., BARBRY P., *An open-access long oligonucleotide microarray resource for analysis of the human and mouse transcriptomes*, Nucleic Acids Research **34**(12), 2006, p. e87.

[7] WATERMAN M.S., *Introduction to Computational Biology. Maps, Sequences and Genom*, Chapman and Hall, London, UK, 1995.

[8] BRESLAUER K.J., FRANK R., BLÖCKER H., MARKY L.A., *Predicting DNA duplex stability from the base sequence*, Proceedings of the National Academy of Sciences of the United States of America **83**(11), 1986, pp. 3746–50.

[9] LONGFELLOW C.E., KIERZEK R., TURNER D.H., *Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides*, Biochemistry **29**(1), 1990, pp. 278–85.

[10] WILSON R.K., MARDIS E.R., *Fluorescence-Based DNA Sequencing, in Genome Analysis: A Laboratory Manual*, Vol. 1, Cold Spring Harbor Laboratory Press, New York 1997, pp. 301–95.

[11] SODA P., RIGON A., AFELTRA A., IANNELLO G., *Automatic acquisition of immunofluorescence images: algorithms and evaluation*, Proceedings of 19th IEEE Symposium on Computer-Based Medical Systems CBMS'06, 2006, pp. 386–90.

[12] CYSEWSKA-SOBUSIAK A., BOLTRUKIEWICZ M., *Modelling of optical signals to be processed for DNA sequence analysis*, Medical and Biological Engineering and Computing **37**(Suppl. 2), 1999, pp. 1148–9.

[13] BŁAŻEWICZ J., CYSEWSKA-SOBUSIAK A., LERCZAK A., KASPRZAK M., MARKIEWICZ W.T., *Parameter analysis of clusters of melting temperatures of DNA chains*, Computational Methods in Science and Technology **3**, 1997, pp. 7–17.

[14] LAKOWICZ J.R., *Principles of Fluorescence Spectroscopy*, Plenum Press, New York 1983.

[15] KELLER P.A., *Electronic display measurement: concepts, techniques, and instrumentation*, John Wiley and Sons, Toronto 1997.
[16] KNOWLES M.B., NHAM T.T., CARTER S.J., *An innovative charge coupled device (CCD) detector for atomic emission spectrometry*, International Laboratory **38**, 1998, pp. 25S–32S.
[17] BOLTRUKIEWICZ M., CYSEWSKA-SOBUSIAK A., PARZYCH J., *Novel approach to modeling of images emitted by a virtual oligonucleotide library*, Proceedings of the SPIE **5959**, 2005, pp. OY1–8.
[18] MCCLUNEY R., *Introduction to Radiometry and Photometry*, Artech House, Boston 1994.
[19] BŁAŻEWICZ J., FORMANOWICZ P., KASPRZAK M., MARKIEWICZ W.T., WĘGLARZ J., *DNA sequencing with positive and negative errors*, Journal of Computational Biology **6**(1), 1999, pp. 113–23.

Received April 25, 2007
in revised form July 11, 2007