UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH
W BYDGOSZCZY

ZESZYTY NAUKOWE
SCIENTIFIC JOURNAL
265

TELEKOMUNIKACJA I ELEKTRONIKA

TELECOMMUNICATIONS AND ELECTRONICS

19

BYDGOSZCZ – 2016

REDAKTOR NACZELNY
prof. dr hab. inż. Józef Flizikowski

REDAKTOR NACZELNY SERII
dr inż. Beata Marciniak

OPRACOWANIE TECHNICZNE
mgr Patrycja Fereni-Morzyńska

Praca powstała przy wsparciu projektu
„Realizacja II etapu Regionalnego Centrum Innowacyjności”
współfinansowanego ze środków Europejskiego Funduszu Rozwoju Regionalnego
w ramach Regionalnego Programu Operacyjnego
Województwa Kujawsko-Pomorskiego na lata 2007-2013

ISSN 1899-0088

Wydawnictwa Uczelniane Uniwersytetu Technologiczno-Przyrodniczego
ul. ks. A. Kordeckiego 20, 85-225 Bydgoszcz, tel. 52 3749482, 3749426
e-mail: wydawucz@utp.edu.pl http://www.wu.utp.edu.pl/

Wyd. I. Nakład 60 egz. Ark. aut. 2,1. Ark. druk. 2,25.
Zakład Małej Poligrafii UTP Bydgoszcz, ul. ks. A. Kordeckiego 20

Contents

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 265
TELEKOMUNIKACJA I ELEKTRONIKA 19 (2016) 5-18

COULD MATLAB HELP TO CURE HIV?

Anna Marciniak[1], Sylwester Kloska[1], Daniel Bujnowski[3],
Vinesh Badloe[2], Elio Abbondanzieri[2], Mahipal Ganji[2]

[1]Nicolaus Copernicus University Ludwik Rydygier Collegium Medicum in Bydgoszcz,
Department of Medicine, Faculty of Biotechnology,
ul. Jagiellońska 13-15, 85-067 Bydgoszcz, Poland

[2]Delft University of Technology, Department of Bionanoscience,
Postbus 5, 2600 AA Delft, the Netherlands

[3]UTP University of Science and Technology,
Faculty of Telecommunications, Computer Science and Electrical Engineering,
al. Prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland

Summary: The human immunodeficiency virus (HIV) is a virus that causes HIV infection and can lead to acquired immunodeficiency syndrome (AIDS). HIV infects cells of the immune system – especially those, which are responsible for the activation of immune response. Every year a huge amount of people die due to diseases that would not be fatal if their immune system was working properly. Scientists from every country want to create an effective drug that helps to cure the infection and prevent development of AIDS. It is necessary to learn everything about HIV to create a drug that will help to save a lot of lives. There is a lot of information discovered by now but also there are some things that remain unknown and should be revealed. One of the most important enzymes for HIV is reverse transcriptase (RT). Thanks to this enzyme virus can re-write its genetic material from RNA (ribonucleic acid) to more stable cDNA (complementary DNA). Finding out the requirements for proper work of RT will help to block and stop the enzyme. A good way to study RT is to observe it under a laser microscope. Laser microscope allows observing single molecules. It is possible to see how RT works with different lengths of DNA (deoxyribonucleic acid) constructs and how does obstacles effect the activity of RT. Results from microscope observations can be analysed using MATLAB software. Special scripts are necessary to analyse binding events and how long they last.

Keywords: HIV, reverse transcriptase, FRET, MATLAB calculations

1. INTRODUCTION

The human immunodeficiency virus (also known as HIV) is responsible for HIV infection and development of acquired immunodeficiency syndrome (AIDS). HIV virus is a lentivirus (subgroup of retroviruses) and contains RNA as genetic material. HIV virus is divided to two types: HIV-1 and HIV-2. Those two types differ between virulence, infectivity and prevalence. HIV-1 type is more virulent and infective than

type HIV-2 and occurs globally, when HIV-2 type occurs mostly in West Africa. Nowadays, HIV is the best known virus, although it is characterized by high volatility, which can manifest itself even in one patient (in various stages of development of the infection). As a retrovirus, HIV virus has to integrate to host genome to replicate itself. To accomplish it, it is necessary to have integrase enzyme and reverse transcriptase (RT). Those two enzymes allow transcription from RNA to cDNA (complementary DNA) and integration to host genome (Fig. 1) [1].

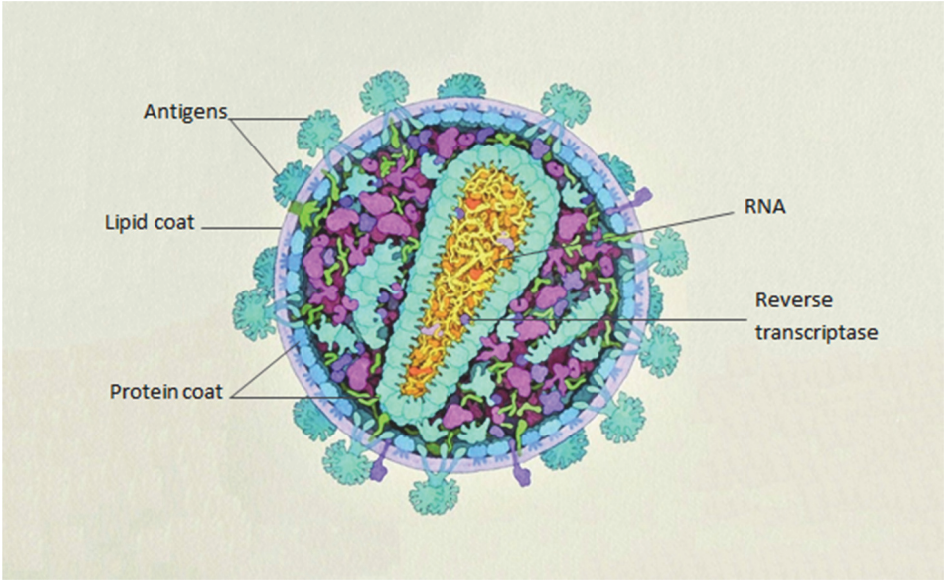

Fig. 1. HIV virus construction – genetic material (RNA) and enzyme (reverse transcriptase) is covered with both protein and lipid coats with antigens, which allows virus to penetrate the host cell [2]

Reverse transcriptase is an enzyme that allows copying single-stranded RNA, which is highly unstable, to more stable double-stranded cDNA. This enzyme is typical for the retrovirus family and some of hepadnaviruses (viruses that contains DNA as genetic material) [1]. RT has 3 functions: it synthesizes DNA on RNA template, synthesizes DNA on DNA template and hydrolysis RNA on DNA template [3]. RT differs between species, mostly in parameters such as molecular weight or number of subunits (Fig. 2). In human cells RT can be found as well. In that case RT is responsible for maintaining length of the telomeres of eukaryotic chromosomes [1, 4].

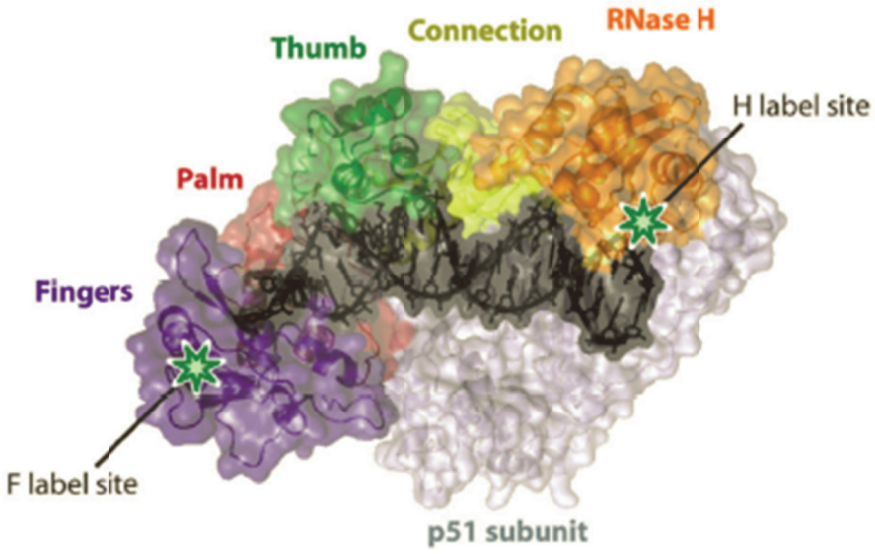

Fig. 2. The structure of HIV-1 reverse transcriptase. Labelling sites for Cy-3 on RT are highlighted by green stars [3]

When virus penetrates the host cell (e.g. human) then its lipid coat disappears and genetic material is released. Then there comes reverse transcription – a process in which the single-stranded RNA is transcribed on cDNA, which is double-stranded on the end of this process. cDNA is capable of integration with genome of the host. This integration plays a key role and is necessary for the next step of virus life cycle. Integrated genetic material uses DNA replication machinery of host to multiply itself (repopulate). When a proper number of copies is attained then the lipid coat is recreated and genetic material is packed inside. New-born particles of virus are being released by the disintegration of a host cell and then they infect other cells of host organism. Thanks to this system, the infection remains dormant for a long time. Cells cannot fight this kind of infection because virus mostly attacks the population of helper lymphocytes (Th), which are responsible for stimulation of immune response. Reverse transcriptase (RT) is an indispensable enzyme for a virus. Thanks to RT virus can replicate itself without any cost because everything that it needs is provided by the host cell [5].

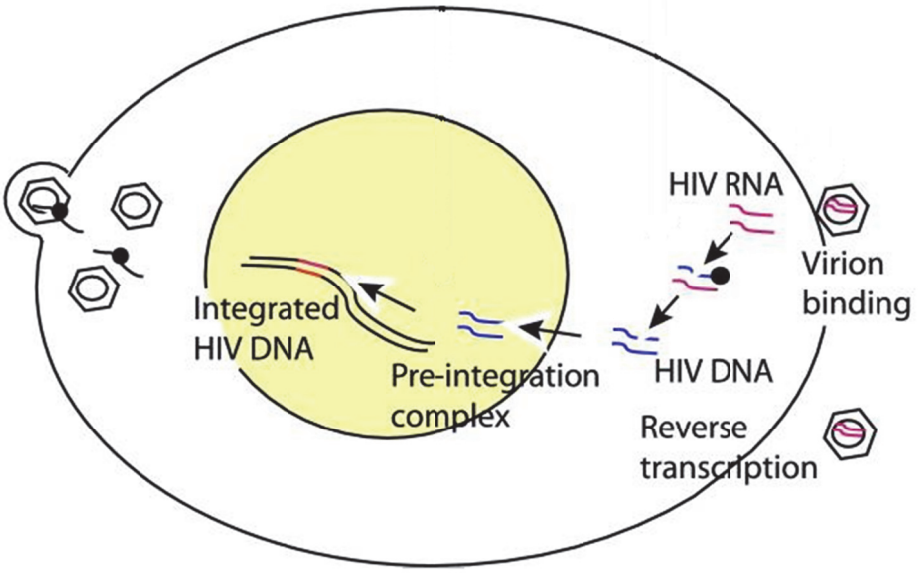

Fig. 3. HIV virus life cycle. HIV virus penetrates host cell. After that, released HIV RNA undergoes reverse transcription process, and then created HIV DNA (cDNA) integrates to host genome and replicates virus particles. In the end new virus particles are released by destroying host cell [5]

AIDS, which is caused by HIV virus was the reason of death in over than one billion people (worldwide) in 2014 [6]. Understanding the mechanism of action of RT may form the basis for the development of effective drugs (or even vaccines) against HIV. Most of the drugs that are used in treating of HIV infection work like inhibitors of reverse transcriptase [7, 8]. They should not allow RT to transcript the genetic material from RNA to cDNA. Unfortunately, so far there have not been invented a drug that would completely stop RT.

In this research we wanted to check total binding time and amount of binding events using new methods of analysis.

2. MATERIALS AND METHODS

The RT, which was used in this experiment, was in-home generated and labelled with fluorophore Cy-3 (cyanine), which maximum of absorption is 532-nm (nanometres). Maximum absorption of Cy-5 is 635-nm. The various length DNA strands (which were a substrate) were labelled with Cy-5, which maximum of absorption is the same as the donor's emission. The beam of 532 nm aroused the labelled enzyme but it did not arouse Cy-5 on the substrate. Cy-5 was aroused only if the donor emitted beam with a proper wavelength. To understand all of this we should take a look on the Jabłoński diagram, which illustrates intramolecular processes of redistribution and excitation energy dissipation of chemical molecule following the absorption of a photon and lead to emissions, i.e. fluorescence or phosphorescence. It allowed the detection of Förster Resonance Energy Transfer (FRET) (Fig. 4) [9, 10].

FRET is a mechanism in which chromophores are capable of transmitting energy, but only if they are close enough to each other (<10 nm). If 2 proteins share a distance of less than 10 nm, it means that the reaction is going on.

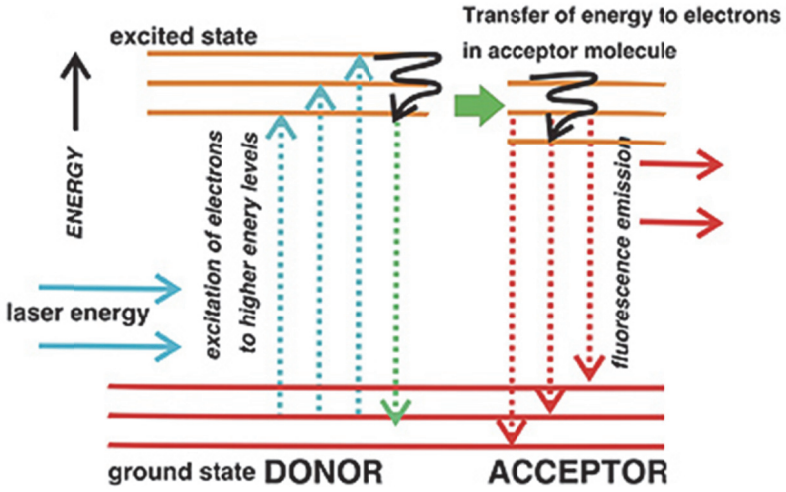

Fig. 4. The scheme of FRET. Laser energy excites donor, then energy is transferred to electrons in acceptor molecule [11]

To observe the FRET we need a laser microscope, which consists of a source of light, mirrors, the beam divider that allows distributing the light on red and green, a prism and CCD camera (charge coupled device). To work a CCD camera needs to be cooled to temperature about -80°C. This temperature is needed because of the speed of taking photos, which can lead to overheating. To watch the results we need a program (e.g. LabView). It is necessary to write proper scripts, which allow alternate switching of green and red laser beams and watching sample on a microscope (Fig. 5).

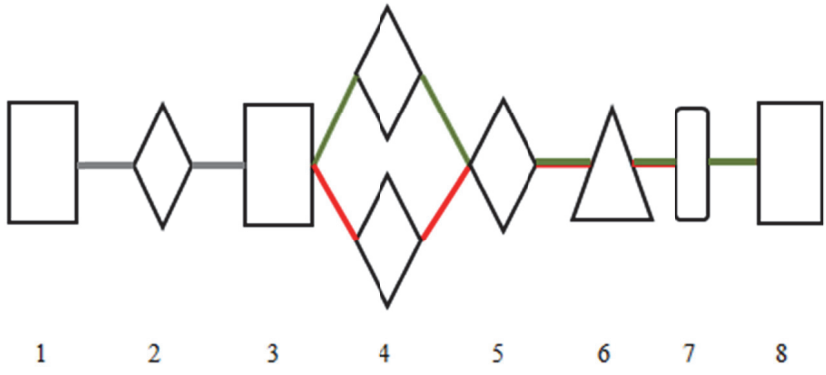

Fig. 5. Scheme of laser microscope. 1 – source of light, 2 – mirror, 3 – separator, 4 – mirrors, 5 – lens, 6 – prism, 7 – glass slide, 8 – CCD camera

To be able to run this experiment it is necessary to prepare a sample in a proper way. Glass slide, which is usually used in microscope observations, was modified this time. To create the flow cells 5 holes was drilled on the longer edges of a slide. A glass slide prepared this way was then covered with polyethylene glycol (PEG) and stored in a temperature of -86ºC. Before being used in an experiment the slide was covered with PEG one more time for about an hour. Then the slide was washed with distilled water and then dried. Double-sided tape was stuck between the drilled holes and after that a cover glass was put on top of it. Last part was to seal the edges with glue to prevent leaking. After the glue has dried, the glass slide was ready for further stages of analysis.

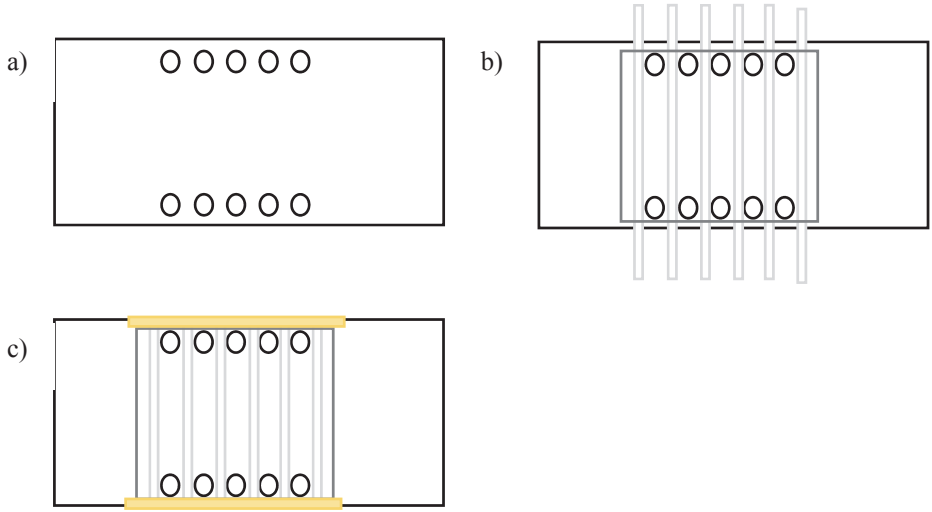

Fig. 6. Glass slide preparation steps: a) drilling holes in glass slide, b) sticking tape and putting on cover glass, c) sealing the edges with glue

In the experiment there was used a linkage between biotin and streptavidin. It is the strongest binding naturally occurring. The strength of this bond is influenced by a few factors, like high complementarity of shapes between so called pocket of streptavidin and biotin. There is also a very extensively network of hydrogen bonds, which stabilizes biotin, when it is in an appropriate position to bind with streptavidin. The "pocket" for biotin has a hydrophobic character (it does not like water). Hydrophobic interactions and the van der Waals bonds, which are also present in there, contribute to the high affinity of biotin and streptavidin. The last thing that should be mentioned in here is an elastic loop, which closes bonded biotin in the "pocket" and thereby contributes to slowing the dissociation. Because of the strength of binding of biotin to streptavidin or avidin, it is used in various fields of molecular biology (during Roche454 sequencing technology), in microbiology and immunology (enzyme-linked immunosorbent assay (ELISA)) to increase the sensitivity of detection [12, 13].

In the experiment the DNA sample was used because of its less demanding requirements of storage and usage. Template was 63 bp (base pairs) long. Primer was 40 nt (nucleotides) long (38 nt complementary and 2 nt non-complementary flap). We tested the activity of RT in various constructs – to described basis and primer we attached constructs in which 15 nucleotides was paired (complementary) and 3, 6, 9, 14

and 23 was not complementary. We also tested the construct without any non-complementary flap. In total there have been 7 constructs tested. We tested the attitude of RT in each of them: the total binding time and the number of binding events.

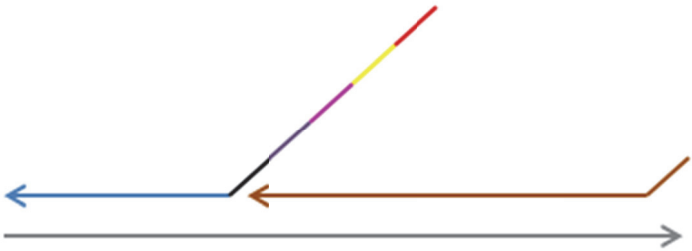

Fig. 7. Scheme of DNA construct. With different colours are marked different flap lengths. The grey colour is 63 bp template, the brown colour is primer (38+2nt). The blue construct is 15 nt long. The additional flaps are 3 nt (black), 6 nt (violet), 9 nt (pink), 14 nt (yellow) and 23 nt (red)

In a first stage of microscopic assay streptavidin was applied to the flow cell and incubated for one minute. After that time the flow cell was washed out with TE buffer plus sodium chloride. Afterwards DNA sample was applied and incubated for one minute. When the incubation was over the last part was to add imaging buffer, consisting of water, 50 mM Tris-HCl buffer (pH = 8.0), 10 nM reverse transcriptase (RT), 1x glucose oxidase (GOx), 0.2 mg/ml bovine serum albumin (BSA), 5% glucose, 0,2% Triton-X100, 100 mM sodium chloride (NaCl), 6 mM magnesium chloride ($MgCl_2$) and 2 mM Trolox. Imaging buffer was applied to add the reverse transcriptase to the flow cell and the components that prevent rapid ordination of fluorophores. Thus prepared, the preparation can be seen under a laser microscope.

Table 1. Components necessary to run microscopic assay. Each of them is added to the solution with specific purpose

Component	**Function**
Water	Dissolvent for other components
Tris-HCl buffer	Maintain proper pH
Various length of DNA strand	Substrate for reverse transcriptase
Reverse transcriptase	Enzyme that carries out process
Glucose oxidase	Used to eliminate oxygen from the reaction
Bovine serum albumin	Used to stabilize enzyme and prevent adhesion
Glucose	Substrate for glucose oxidase
Triton x100	Detergent, reduce surface tension
Sodium chloride (NaCl)	Different concentration of this component allows modulating reaction speed
Magnesium chloride ($MgCl_2$)	Supplies Mg^{2+} ions for proper working of RT
Trolox	Antioxidant, protects DNA sample from damage

The preparation was placed under the laser microscope. Properly written scripts in LabView software allowed alternately irradiation with the green and red laser and recording the movie. A highly sensitive, monochrome CCD digital camera was used to save images. It used the active cooling system. This type of cameras is dedicated to work in the field of cell biology that requires short exposure times (fluorescence). Movie consisted of a series of photos made with CCD camera each 0.1 sec for around 210 sec. To correlate the molecules that were irradiated with various laser colours, a mapping function was used. Mapping allowed to match and merger molecules from two screens (Fig. 9). Then there was selected area, from which molecules were chosen for the next stages of analysis. For this purpose was used ImageJ software (Image Processing and Analysis in Java) [14]. Molecules that were on edges have been removed from further analysis to avoid false results. ImageJ software chooses the local maxima depending on determined noise tolerance. Position of these molecules was presented in Cartesian – each molecule had X and Y coordinate (Fig. 10). Then those coordinates was saved as a list in "*.txt" file. The next stage was extraction of the data in MATLAB software.

```
% Import data from file 'txt'
PeaksFileName = dir('*.txt');
Peak=importdata(PeaksFileName(1,1).name);
Cy5Xcoordinate=Peak.data(:,2)+1;
Cy5ycoordinate=Peak.data(:,3)+1;
```

Listing 1 Load values for the coordinates of the txt file

Data are processed in three MATLAB files. First of them is responsible for pre-processing data. Listing 1 shows only the most important lines of code, how to import data for appropriate channels.

Afterwards it is necessary to obtain coordinates in green channel. Coordinates are calculated from mapping and peaks of the red channel. Code showed on Linsting 2 is responsible for create three matrixes needed to obtain coordinates in green channel. In two of them (p_r_x & p_r_y) are stored data corresponding to coordinates in red channel. In 3rd matrix are calculated data to 4th order polynomial.

```
% Obtain the coordinates in green channel from mapping and
peaks from red channel
for i=1:length(Cy5Xcoordinate)
  if Cy5Xcoordinate(i)>0
    p_r_x(k)=Cy5Xcoordinate(i);
    p_r_y(k)=Cy5ycoordinate(i)-256;
    %4th order polynomial
  args(k,:)=[1 p_r_x(k) p_r_y(k) p_r_x(k)*p_r_y(k) p_r_x(k)^2
p_r_y(k)^2 p_r_y(k)*p_r_x(k)^2 p_r_x(k)*p_r_y(k)^2 p_r_x(k)^3
p_r_y(k)^3 p_r_x(k)^3*p_r_y(k) p_r_x(k)^2*p_r_y(k)^2
p_r_x(k)*p_r_y(k)^3 p_r_x(k)^4 p_r_y(k)^4];
    k=k+1;
  end
end
```

Listing 2 Matrixes storing data for the green channel

Then in the decision process are calculated coordinates for the green channel, depending on the level of the red channel (listing 3).

```
for i=1:length(args(:,1))
    %the transformation type is polynamial
    if p_r_x(1,i)<=140
        p_g_x(i,:)=args(i,:)*mytform1.tdata(:,1);
        p_g_y(i,:)=args(i,:)*mytform1.tdata(:,2);
    elseif p_r_x(1,i)>140 & p_r_x(1,i)<=370
        p_g_x(i,:)=args(i,:)*mytform2.tdata(:,1);
        p_g_y(i,:)=args(i,:)*mytform2.tdata(:,2);
    elseif p_r_x(1,i)>370
        p_g_x(i,:)=args(i,:)*mytform3.tdata(:,1);
        p_g_y(i,:)=args(i,:)*mytform3.tdata(:,2);
    end
    end
```

Listing 3 Calculating data for the green channel

Next the data are aggregated and regions of interest are saved in to file (listing 4)

```
FnamaGreen=[fname 'GreenRegionsOfInterest' num2str(l) '.mat'];
save(FnamaGreen,'GreenRegionOfInterest','-mat') % saves all
the extracted green regions of interests
FnamaRed=[fname 'RedRegionsOfInterest' num2str(l) '.mat'];
save(FnamaRed,'RedRegionOfInterest','-mat')% saves all the
extracted red regions of interests
```

Listing 4 Save the selected regions to files

Next file contain the code which is reduce level of noise in an image. Algorithm import pre-prepared data and process them to reduce level of noise. Code shown on listing 5 is responsible for calculate level of threshold which is used to decide is whether the data or noise. The process of reduce noise in green channel is performed similarly.

```
%Detecting the threshold for red trace background correction
    STR=sort(tr_r);

pri=polyfit(1:round(3/4*length(tr_r)),STR(1:round(3/4*length(t
r_r)))',1);

pre=polyfit(1+round(3/4*length(tr_r)):length(tr_r),STR(1+round
(3/4*length(tr_r)):length(tr_r))',1);
    RedThreInd=round((pre(2)-pri(2))/(pri(1)-pre(1)));
    if RedThreInd<numel(STR)&& RedThreInd>0
        RedThre=STR(RedThreInd);
    else% RedThre>150
        RedThre=35;
    end
```

Listing 5 Process to reduce noise in red channel

Last step is to display all processed data on plot to compare and analyse them. Responsible for this is code shown on listing 6. Presented code is used to set limit of axis, colours of bar, method display data etc.

```
figure()
h=bar(g,prob,1,'b','EdgeColor','k');
set(gca,'FontSize',18,'LineWidth',3)
set(h,'FaceColor',[0.75 0.75 0])
xlim([-0.20 1.25]) %set limit on X axis

hold on
[fo,gof]=fit(g',prob','gauss2','startpoint',[0.025, 0.25, 0.2,
0.021,1,0.4])
hold on
plot(-0.2:0.01:1.3,fo(-0.2:0.01:1.3)','color',[0 0 0],
'linewidth',2.5)
xlabel('FRET','fontweight','b','fontsize',22)
ylabel('Fraction','fontweight','b','fontsize',22)
title('63/38+2/15nt DNA', 'fontweight','b','fontsize',16)
RatioOfFretAreas=sum(fo(-0.2:0.005:0.68))/
sum(fo(0.68:0.005:1.2))
```

Listing 6 Chart of analysed data

During this stage the results of a single molecule was saved. We observed if there occurred the binding of RT with DNA construct (Fig. 8). The selection of molecules was made manually. The DNA construct was labelled with Cy-5. The red colour was showing the signal emitted by excited with the laser dye Cy-5. The green fluorescence comes from the reverse transcriptase which was labelled with Cy-3. There is a binding event if on the analysed image is a significant (distinctive from the background) increase of the red signal or increase of both red and green signals simultaneously. If there is increase of the green fluorescence without red signal, there is no binding event – these signals should be removed with the help of proper scripts in MATLAB.[3] On this stage it is important to remove the background „noise" and false results, which were generated by RT that was not tied with the DNA. After removing of all noises, there are only binding events on the screen. Then the results were subjected to statistical analysis (Gaussian distribution). Thanks to the results it is possible to learn e.g.: how many binding events is in a single molecule during the checked time, how long they last and it is possible to compare the activity of RT in different conditions by changing the reaction mix (for example the concentration of sodium chloride (NaCl)).

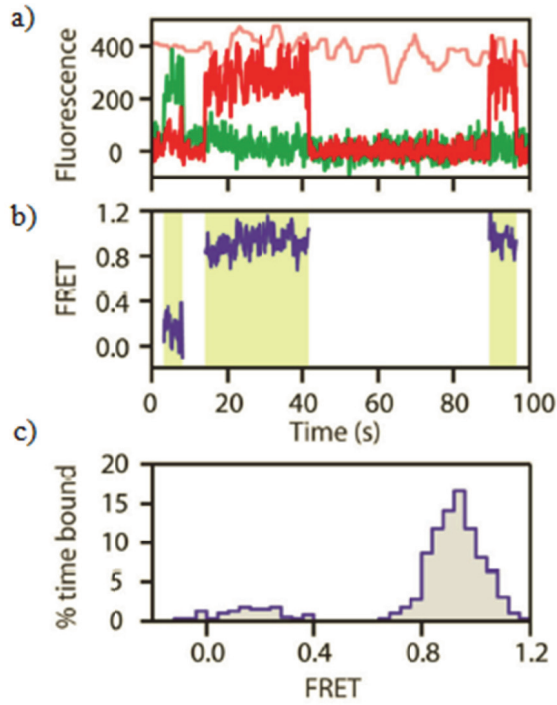

Fig. 8. The results of analysis. a) Increase of red signal shows a binding of RT and DNA. b) Noises were removed from the background, binding time and FRET values are shown. c) FRET distribution histogram [3]

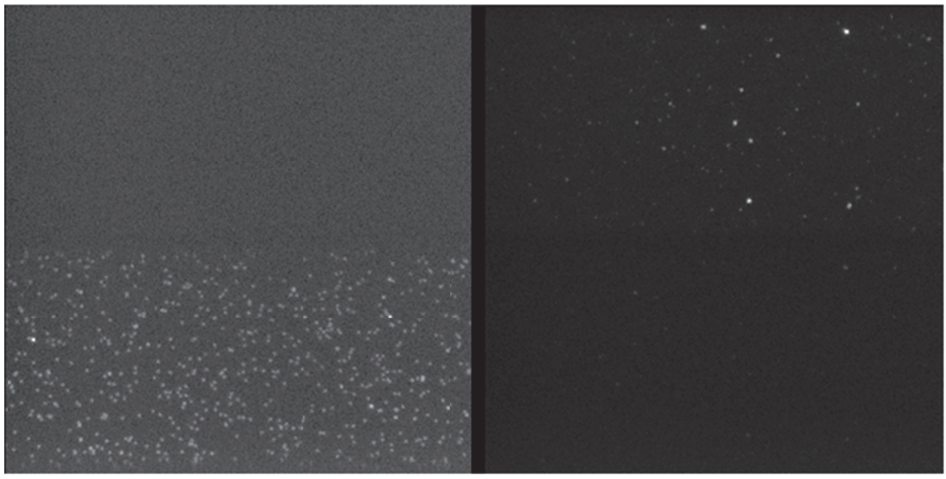

Fig. 9. Image obtained from the laser microscope. On the left side is shown result of red laser irradiation (DNA construct). On the right side is shown result of green laser irradiation (reverse transcriptase particles)

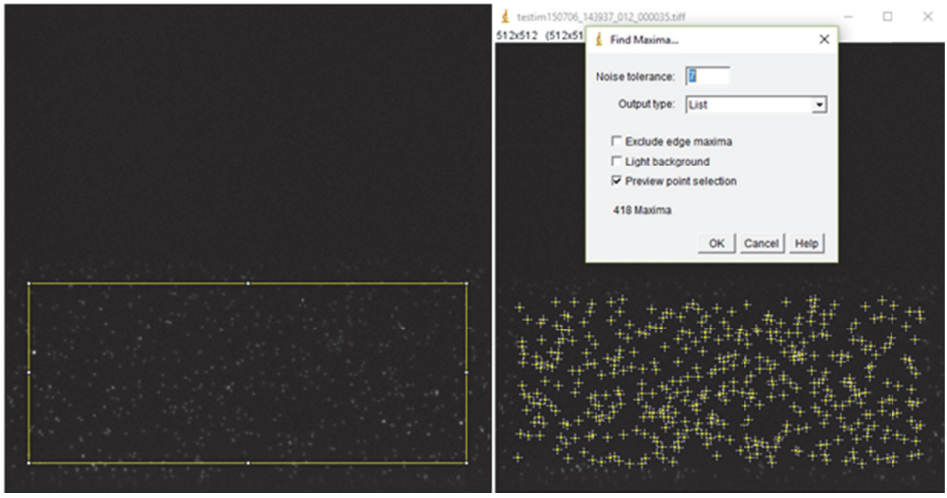

Fig. 10. ImageJ analysis. On the left side is shown clipping area for further analysis. On the right side is shown creating list of molecule coordinates

3. RESULTS

In this experiment we obtained information about number and time of binding RT to DNA construct. This data can be used to calculate chemical parameter value, such as dissociation constant (K_d) or statistic values such as standard deviation (e.g. from K_d) and Gaussian distribution (e.g. of number of binding).

4. CONCLUSIONS

Understanding of reverse transcriptase is a crucial way to discover a new way to treat HIV infections. The mortality of HIV infection is very high and the ways to treat that kind of infection are not good enough. Described method of analysis works for testing the activity of reverse transcriptase in different DNA constructs. It can be helpful in finding new ways to stop RT activity. To carry out the analysis in a way described in this paper, it is necessary to be familiar with LabView and MATLAB software. Properly written scripts allow performing correct analysis, obtaining trustful results and process automation. Reproducible results are necessary to draw appropriate conclusions. One of the ways it can be accomplished is the common work of biologists and IT specialists. To find new ways for successful analysis it is necessary to combine those two science disciplines. Merger of biology and computer sciences, so called bioinformatics, is the future of nature sciences and medicine.

BIBLIOGRAPHY

[1] Hu W.S., Hughes S.H., 2012. HIV-1 reverse transcription, Cold Spring Harb Perspect Med. 2, 1-22.
[2] Johnson G.T., Goodsell D.S., Autin L., Forli S., Sanner M.F., Olson A.J., 2014. 3D molecular models of whole HIV-1 virions generated with cellPACK., Faraday Discuss. 169, 23-44.
[3] Abbondanzieri E.A., Bokinsky G., Rausch J.W., Zhang J.X., Le Grice F.J., Zhuang X., 2009. Dynamic binding orientations direct activity of HIV reverse transcriptase, Cancer. 453, 184-189.
[4] Stone D.M., Mihalusova M., O'Connor M.C., Prathapam R., Collins K., Zhuang X., 2007. Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein, Nature. 446, 458-461.
[5] Murray J.M., Kelleher A.D., Cooper D.A., 2011. Timing of the components of the HIV life cycle in productively infected CD4+ T cells in a population of HIV-infected individuals., J. Virol. 85, 10798-10805.
[6] WHO - HIV Department, Global summary of the AIDS epidemic, (2014). http://www.who.int/hiv/data/epi_core_july2015.png?ua=1.
[7] Das K., Arnold E., 2013. HIV-1 reverse transcriptase and antiviral drug resistance. Part 1., Curr. Opin. Virol. 3 (2013) 111-118.
[8] Das K., Arnold E. HIV-1 reverse transcriptase and antiviral drug resistance. Part 2., Curr. Opin. Virol. 3, 119-128.
[9] Myong S., Bruno M.M., Pyle M.A., Ha T., 2007. pring-Loaded Mechanism of DNA Unwinding by Heptitis C Virus NS3 Helicase, Science. 317, 513-516.
[10] Kupfer S.S., Torres J.B., Hooker S., Anderson J.R., Skol A.D., Ellis N.A., et al., 2009. Novel single nucleotide polymorphism associations with colorectal cancer on chromosome 8q24 in African and European Americans., Carcinogenesis. 30, 1353-1357.
[11] Bio-Imaging Unit, FRET, (n.d.) http://www.ncl.ac.uk/bioimaging/techniques/fret/.
[12] Diamandis E.P., Christopoulos T.K., 1991. The biotin-(strept)avidin system: Principles and applications in biotechnology, Clin. Chem. 37, 625-636.
[13] Frampas E., Rousseau C., Bodet-Milin C., Barbet J., Chatal J.-F., Kraeber-Bodéré F., 2013. Improvement of radioimmunotherapy using pretargeting., Front. Oncol. 3, 159.
[14] ImageJ. http://imagej.nih.gov/ij/.

CZY MATLAB MOŻE POMÓC WYLECZYĆ HIV?

Streszczenie

Infekcja wywołana ludzkim wirusem niedoboru odporności (HIV) może prowadzić do zespołu nabytego niedoboru odporności (AIDS). Wirus HIV infekuje komórki układu odpornościowego – zwłaszcza te, które odpowiedzialne są za aktywację odpowiedzi immunologicznej. Każdego roku wiele osób umiera z powodu chorób, które w wypadku prawidłowego działania układu odpornościowego nie byłyby śmiertelne. Aktualnie naukowcy próbują opracować skuteczny lek, który pomoże leczyć infekcję wirusem HIV i będzie zapobiegać rozwojowi AIDS. Aby to osiągnąć konieczne jest jak najlepsze poznanie

cząsteczki wirusa HIV i sposobu jego działania. Do dnia dzisiejszego odkryto wiele informacji o wirusie HIV, jednak wiele jego właściwości pozostaje nieznane**.** Jednym z niezbędnych enzymów wirusa HIV jest odwrotna transkryptaza (RT). Dzięki temu enzymowi wirus może przepisać swój materiał genetyczny z RNA na bardziej stabilne cDNA (ang. complementary DNA). Poznanie warunków, w których działa odwrotna transkryptaza pomoże zablokować jej aktywność. Dobrym sposobem na poznanie tego enzymu jest jego obserwacja pod mikroskopem laserowym. Mikroskop laserowy umożliwia obserwację pojedynczych cząstek. Możliwa staje się obserwacja reakcji RT z konstruktami DNA o różnej długości. Wyniki otrzymane z obserwacji pod mikroskopem mogą być analizowane za pomocą programu MATLAB. W tym celu konieczne jest napisanie odpowiednich skryptów, które pozwolą na dokładną analizę aktywności odwrotnej transkryptazy.

Słowa kluczowe: HIV, odwrotna transkryptaza, FRET, obliczenia w MATLABie

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 265
TELEKOMUNIKACJA I ELEKTRONIKA 19 (2016) 19-28

SECURITY IMPROVEMENT IN A MOBILE PAYMENT SYSTEM

Mścisław Śrutek[1], Agata Wojciechowska[1], Josep Solé-Pareta[2]

[1]UTP University of Science and Technology,
Faculty of Telecommunications, Computer Science and Electrical Engineering,
al. Prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland

[2]Universitat Politècnica de Catalunya (UPC), C. Jordi Girona, 31
08034 Barcelona, Spain

Summary: The mobile payment system and possible ways of using it are presented in this paper. There are a security analysis and a description of a potential risk. A proposal of security improvement is also included in the paper. The proposed solutions may be both safe and comfortable for mobile payment users. This paper is based on the research done as part of the COLIBRI Erasmus+ program and available online documents.

Keywords: mobile payment system, security, COLBRI Erasmus+

1. INTRODUCTION

Modern technologies have been progressively introduced into the people life's. Hardly anybody has thought about children, who readily use tablets, or about the elderly people spending their free time on the Internet. SmartWatch has become a transparent standard combined with the phone function, so has Smart TVs with higher resolution matrices – with access to the Internet, as well as the phones having more computing power than any computer in the past. In light of this progress, revolution in the banking sector is a natural consequence.

Such changes could not occur without participation of scientists from technical universities. In order to study the Future Internet Opportunities, a COLIBRI course has been established as part of the European Erasmus + program [2]. The project includes 7 universities from 7 countries: Denmark (Aalborg University), Norway (University of Stavanger), Latvia (Riga Technical University), Germany (Technical University of Hamburg), Poland (University of Science and Technology in Bydgoszcz), Turkey (Bogazici University in Istanbul) and Spain (Technical University of Catalonia in Barcelona) and three business representatives: Atene mobile in Berlin, Talaia Networks in Barcelona, EKT / NHRF in Athens. The overall objective of the project is work in cross-cultural and cross-disciplinary group upon use of the latest technologies in the field of IT, as well as anticipating future developments and solutions. Some of the ongoing issues relate to economics and entrepreneurship. Moreover, for academic teachers this is an opportunity to get familiar with new learning methods and confront their experiences with the knowledge of lecturers from other countries. The participants

of the course are students of the above mentioned universities (3-5 students per university). Together they carry out courses on various topics and participate in workshops (including video lectures, assignments and activities covering the future Internet from different points of view). They are divided into smaller international groups and under the guidance of lecturers they implement various projects according to the latest industrial trends. The themes of the projects cover real problems reported by companies operating on the markets of different countries. One of the projects, implemented in the 2015 course of COLIBRI was '*The personalization vs. privacy tradeoff in a mobile-payment experience*'. The project was carried out on behalf of DINUBE[3], a company from the mobile banking sector in Spain. Dinube was asked for reporting the expectations of the users, and their needs, while the company was interested in getting to know how those expectations should be fulfilled in the best possible way, using the most innovative technology. This knowledge would enable DINUBE to provide more comprehensive services, by increasing mutual trust, thus increasing the number of customers.

Within the scope of this project, online survey was carried out, in order to indicate the opinion of the Internet users on key security issues. As a result, more than 200 respondents from seven countries of Europe provided their answers to the questions of the survey. The survey included, among others, questions about reading the conditions of privacy policy and regular changes of passwords.

The article provides some results of the considered survey and analyses current solutions in the field of mobile payments. Some of the most popular applications and their most important features have been taken into consideration. As a result, an innovative solution has been offered, that can improve the security of mobile payments while maintaining the convenience of use.

The mobile payment market is relatively new and is changing rapidly. For this reason in literature there are only sets of online papers. This set was completed carefully to rely only on very reliable sources, e.g. European Central Bank [4] or the financial branch leader Visa [10]. In this paper there are also references to annual reports. They concern the usage of smartphones, modern technologies and development prospects [7, 8]. In literature there are also links to the Security Research Labs's documents that refer to breaking security measures [5, 9] and link to the COLIBRI course home site [2].

2. MOBILE PAYMENT ANALYSIS

Nowadays each or almost each Pole has a mobile phone. The majority of them (over 60%) uses smartphones [8]. According to the telecommunication companies, the sale of basic phones is constantly decreasing. They try to convince their clients to the smartphones but at the same time they do not withdraw the basic phones from sale. The smartphone possession is closely related to the age and is presented in the Figure 1.

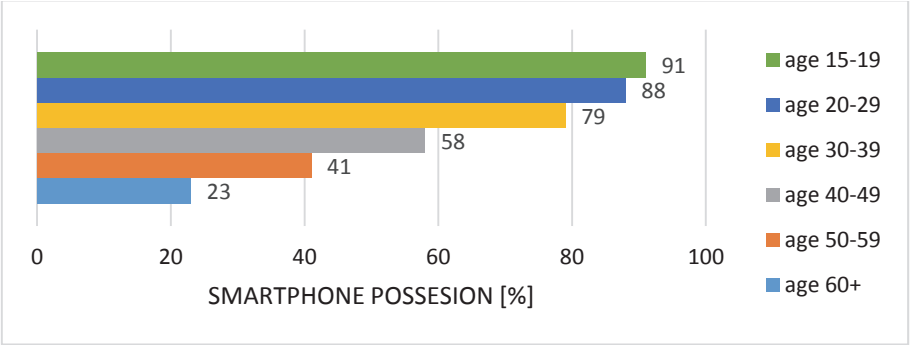

Fig. 1. The smartphone possession in Poland in 2015[8]

Mobile banking is rather a young service. In Poland the first attempt to implement banking operations into the mobile world was in 2000. Because of high prices of data transmission, this service did not belong to the mainstream. Rapid development of the mobile payment system started together with reduction of data transmission prices. Banks and other institutions from the banking sector have been creating and publishing their own mobile applications for about 5 years. These applications are catching users' attention and are constantly changing the clients' attitude to the mobile payment. The figures presented below show how the attitude to the mobile payment has changed since 2013.

Mobile payment has become increasingly popular for the last two years. For this reason most banks provide their clients with a possibility to use a specific mobile application with a wide range of features. There are some features listed in the table below, but the most basic functions including: checking the bank account balance, using bank transfers or paying at a different kind of shops, have been intentionally neglected. They are simple and each mobile payment application can realize them.

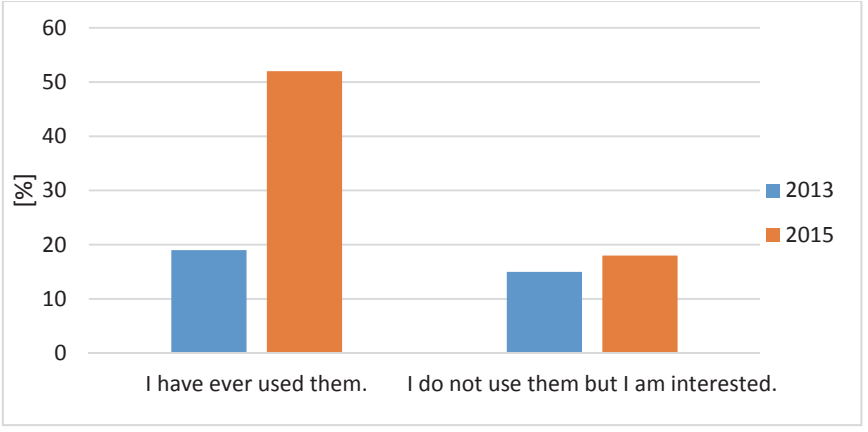

Fig. 2. The changing attitude to mobile payment in Poland from 2013 to 2015 [7, 8]

It is worth seeing that the mobile applications are dedicated to the different kinds of mobile operating systems. Software development focus mainly on the most popular platforms (Android) but still there are some applications for niche operation systems (BlackBerry).

Table 1. Mobile payment applications' functions [6]

Bank	Bank machine	P2P payment	BLIK	Prepaid mobile phone	Operating systems
Bank Pekao	yes	no	no	yes	Android, iOS, Windows Phone, BlackBerry, Symbian
ING Bank Śląski	yes	yes	yes	yes	Android, iOS, Windows Phone, BlackBerry
PKO BP	yes	yes	yes	yes	Android, iOS, Windows Phone, BlackBerry, Symbian
BZ WBK	yes	yes	yes	yes	Android, iOS, Windows Phone
Bank Millenium	yes	yes	yes	yes	Android, iOS, Windows Phone
Alior Bank	yes	yes	yes	yes	Android, iOS, Windows Phone
mBank	yes	yes	yes	yes	Android, iOS, Windows Phone
Eurobank	no	no	no	yes	Android, iOS, Windows Phone
Getin Bank	no	no	no	no	Android, iOS, Windows Phone
Bank BPH	no	no	no	yes	Android, iOS

The column especially worth seeing in the table is the 'BLIK' column. It refers to an additional service that is included in mobile applications form 6 banks (another banking companies will be joining this program over the next months) and is called BLIK [1]. This is the service delivered by Polski Standard Płatności (en. Polish Payment Standard, shortly named as PSP). It has been prepared since 2013, until on 9th of February 2015 it was officially started. In December 2015 there were over 1.5 million of users and over the million transactions done with BLIK. The main features include:

- payment in shops and service points,
- payment online,
- withdrawing cash from banking machine,
- bank transfers using only the telephone number of a recipient.

Moreover there is a special loyalty program for the BLIK users, they may buy cheaper cinema tickets or may have a lower price for VOD movies. Even though the number of the service points that accept this kind of payment is constantly getting bigger, it is still a solution available only on the internal Polish market.

Another way of mobile payment is the HCE (Host Card Emulation) technology using. The most distinctive features of this kind of solution are making use of the NFC (Near Field Communication) module and moving all needed computing into a cloud. Moreover, the owner of a smartphone is not bound to exchange his/her SIM card in order to make the HCE payment possible. The application needed to run the operation is uploaded on the terminal.

When the HCE payment is being performed, a smartphone operates as a common proximity card that may be used to do any contactless operation. HCE operations are available for clients of the Polish banks such as: Getin Bank, BZ WBK or Pekao. However, the requirement for the smartphone to use the HCE payment are Android operating system (version 4.4 KitKat or higher) and NFC module included in the mobile handset. The biggest advantage of this solution is its wide acceptance. By the end of 2017 it will be possible to use the HCE payment in each terminal in Poland while by the end of 2019 in each terminal in Europe. This kind of payment is supported by Visa and MasterCard and thanks to this support it is possible to pay with HCE even in places without access to the mobile network.

However there is still a possibility to use mobile payment with an older type of smartphone (without NFC module) or with an operating system different from Android. This possibility is based on the QR codes. The QR codes are commonly used to keep static information about a bank transfer. They are mostly placed on the invoices coming from mobile operators or electricity suppliers. They include the basic information about the transfer, like a recipient, a topic of the transfer, an amount of money that should be paid. There is another usage of QR codes in mobile payment applications coming from banks. The code is automatically generated in the terminal, users scan this code with their mobile phones and accept the started transaction with their personal PIN numbers.

3. CURRENT SOLUTIONS AND THEIR SECURITY LEVEL

The anonymous online questionnaire was published in order to discover what users' requirements are, and which factors may make the respondents become a mobile payment system users. This survey was done as a part of the COLIBRI program. There were more than 200 answers collected in the questionnaire. However, most of respondents was students of technical universities or engineers. They may concern more about the technological issues. In the former analysis, the survey should be targeted also to other groups of general population.

The figure presented below shows the respondents' answers to the question 'What attributes are strong incentives for you to use mobile payment?'

It comes from the figure that according to the users' answers the most important issue is the privacy (61%). It may be connected somehow with the news published on the Internet and by the press telling about some data leaks. In some cases, sensitive data leaked out from servers of different companies and became public. That may be the reason why respondents are worried about their personal data.

The figure shows also that, according to the users' declarations, they pay a lot of attention to the security (44%). The mobile payment concerns financial issues and losing a big amount of money may cause some serious consequences. Hence, the software companies should pay more attention to the security of the mobile payment. The applications have to be secured from unauthorised access and the security should be absolutely reliable.

However, there is still a relatively big number of users who really do not attach much importance to keeping their private data safe. They appreciate more the possibility of using the mobile payment than security. The great challenge for designers of applications is to take into account both of these requirements.

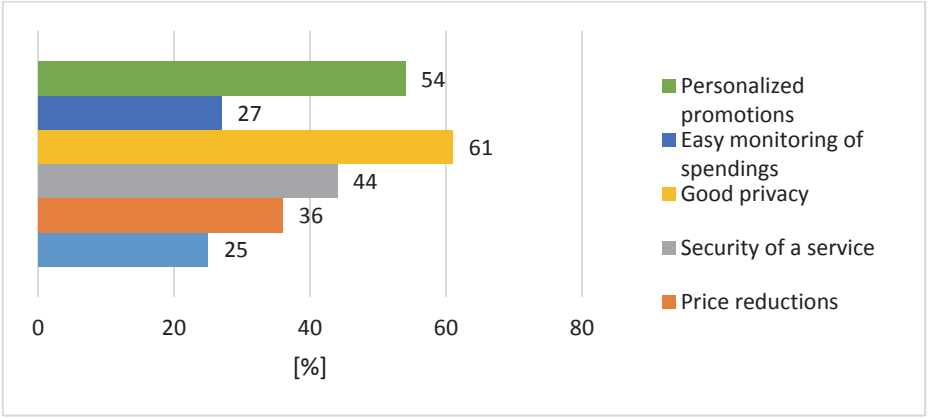

Fig. 3. The online questionnaire's results

In some further paragraphs different ways of security are presented. They refer to the currently used solutions starting from the BLIK [1] system and finishing with data localization. The BLIK system is described as fast, simple and safe. The main authorisation way is a special code. Each code is generated by the PSP as a chain of six random digits. It is valid only for two minutes from the moment of generation. In addition, in order to generate a code, the user has to log in to the mobile payment application which is basically secured with the personal PIN code. The process of a code authorisation in the BLIK system has five steps:

- the code generation,
- putting the code into the terminal,
- the code authorisation by PSP,
- the operation authorisation by the bank,
- transferring an answer to the store.

There are also some possibilities to use biometrical data. Some banking applications may be authorized with the user's fingerprint. This opportunity is given to the clients of banks Millenium, ING and mBank (a service available only for corporation clients). However, in order to enable a biometrical authorisation a user has to possess a selected model of the smartphone, there are for the iOS operating system: iPhone s5, iPhone 6, iPhone 6 Plus and for the Android operating system, there are three Samsung's devices with a special Samsung Pass function available only on the newest mobiles: Galaxy S5, Note 4 and Galaxy S6. Apart from this biometric authorisation it is still possible to log in with a standard PIN code that is composed of four digits.

The security in the HCE payment system is similar to the proximity card security. If the NFC module is active it enables payment right after unlocking the screen and approaching it to the terminal. When the user does the transaction (e.g. shopping) for the amount of money lower than 50 PLN, he/she will not be asked for PIN code. In case of prices higher than 50 PLN, the user will have to enter his/her PIN into the terminal.

There are not many operations that can be performed without any authentication in the mobile payment applications. One of them is checking the bank account balance. It is worth emphasizing that it is the most common operation in mobile banking. It is much faster as the user does not have to enter the password and wait for an

authentication. There is a significant difference between the current solution and former ones such as plastic debit card with a small display in the corner (introduced by Getin Bank in 2013). In case of cards, two PIN codes were used in the past. The first one was used only for checking the bank account balance while the second was used for an operation authentication. The majority of mobile applications (like mBank, ING, Millenium) do not show the account's balance directly. At the beginning, the user has to set the maximum balance level and afterwards only the percentage is visible. It may improve the security because nobody knows how big the maximum level is. However, there are still applications (BZ WBK) that without logging in show the account balance in PLN.

Some mobile payment applications use localisation data. After the user's acceptance, the application may analyse his/her position and show the nearest bank agency. In case of the BLIK system, the user may receive the full information about the nearest cash machines, shops and service points that accept payment with a BLIK code.

4. PROPOSAL OF SECURITY IMPROVEMENT

The statement saying that any solution may provide complete safety of the system cannot be true. There is always the element which may break and damage the whole security. Unfortunately, the users seem to be the weakest part of a security system. People are able to remember a countable number of logins, passwords and numeric codes. Furthermore, the knowledge about the potential risk connected with the Internet and the newest technology is decreasing with the user's age.

Some questions about the user's behaviour online was answered in the COLIBRI's questionnaire (Fig. 4, Fig. 5). It was intentional to ask about real habits not about the rule.

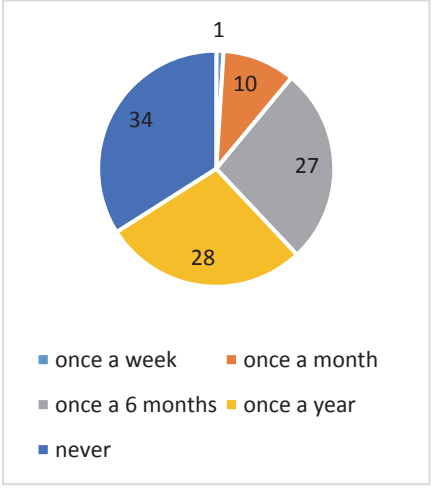

Fig. 4. COLIBRI questionnaire: 'How often do you change your passwords?'

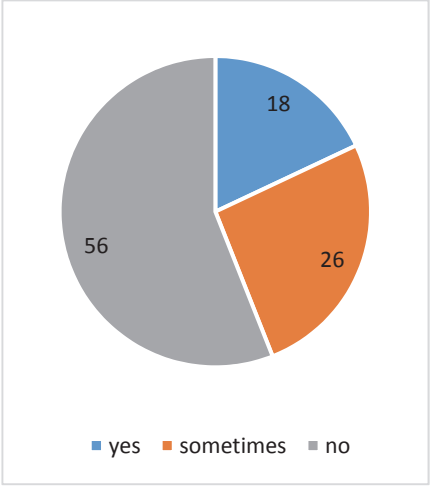

Fig. 5. COLIBRI questionnaire: 'Do you note your passwords anywhere?'

Results coming from two questions were presented in the figures. It is easy to see that one in three users does not change his/her passwords to applications and accounts at all. However, positive is the fact that more than half of respondents do not write down their passwords and codes anywhere.

Unfortunately, current securities sometimes are not good enough to protect the mobile applications from frauds. In further presented paragraphs there are some biggest doubts about each commonly used technology. First of all, there is a PIN code or a password. This is the most basic way of protection of the application. The default code length is four digits which gives about 10.000 potentially solutions. Obviously, the security increases with the password length. Nevertheless, it is possible for a potential thief to check all the possibilities with the current computing power and it would be not very time consuming. Moreover, codes and passwords defined by users are usually relatively short and schematic (e.g. the sequence composed of the same digit or the sequence similar to 12345). People sometimes use bank machines or terminals carelessly. They do not cover the keyboard anyway so that everyone can easily see their PIN code. They do not realize that there may be a small camera installed above the bank machine's keyboard. The next mistake may be writing down the passwords on the small pieces of paper or in case of a mobile payment in the telephone notes.

The HCE payment system is provides security similarly to the proximity card system. Any transaction below 50 PLN is authorised without any confirmation. Hence, in case of stealing of the mobile phone a thief has full access to the user's money. Obviously, transactions may be insured but it depends on the insurance conditions which are described in the agreement in details. There is also a possibility for user to reduce the limit to 0 PLN. In this way the user has to enter the PIN code each time he/she uses the application.

The next way of protection of the mobile payment application is biometrical data. It should be remembered that the fingerprints scanner that are mounted into the mobile devices might be easy to deceive. The Security Research Labs [5, 9], the group of German scientists, proved that this protection can be easily broken. Each person leaves their fingerprints on many places like: a smartphone screen, a computer keyboard, a desktop or even a door. German scientists used the unintentionally left fingerprints, put them on the special foil and prepared the pattern. They were able to unlock the Samsung Galaxy S5 and iPhone 5s by swiping this pattern through the scanner. Moreover, there were some information available on the Internet warning about a possibly wrong storage policy. BMP files with fingerprints are probably stored on HTC and Samsung devices without having proper security.

The last two types of protection are a point unlock (well-known from Android smartphones joining dots in the proper order) and face recognition. Using the point unlock is similar to PIN codes and passwords. This gesture may be seen by an unauthorised person. The second security is not able to detect if there is a real human in front of the camera or maybe there is only his/her photography. According to the above mentioned analysis, the most important threats include:

a) watching by an unauthorised person (a PIN code),
b) counted number of combinations (a pattern unlock),
c) sensitive data storage inside the device's memory (fingerprints),
d) using a fake pattern (a face recognition).

A gesture recognition seems to deprived of these mistakes. The gesture should be natural and possibly done many times during the day, like: smile, eye wink. The software should not compare the user's face with the remembered pattern (no data stored) but only analyse the movement and detect the defined gestures. It may be reasonable to put the gestures in a sequence. In that case only the sequence of proper gestures gives the authorisation to the mobile transaction.

5. CONCLUSIONS

Modern technologies are a great chance for the bank's clients. Payments are getting much faster and more comfortable. Unfortunately, users often forget about the proper security of their data and money. They sometimes prefer more comfortable services than safe services. Most mobile payment systems are presented in this document including their main features and, above all, their potential disadvantages.

Obviously, it should be emphasized that payments done with the mobile phone are not bad or useless. They are really modern, still in progress and comfortable. The application user knows his/her bank account balance and can easily manage his/her expenses. The security of mobile payment applications has to be the priority for banks. Nevertheless, the weakest part of the authorisation chain is the user. The proposed way of protection using a gesture recognition gives the user the maximum of comfort without the risk of losing money.

Now the main aim for banks and other companies from the financial sector should be the education. They should persuade the users that by obeying the rules (not writing down the passwords, changing them regularly etc.), users may help in the mobile payment development. Education and further development of security issues will definitely increase the number of the mobile payment system users.

The research on the mobile banking started at COLIBRI course will be continued. The improvement proposed in this article will be soon implemented and used to create the master project.

BIBLIOGRAPHY

[1] BLIK home web page, access: [01.2016 www.blikmobile.pl].
[2] COLIBRI project web page, access: [01.2016 http://www.tuhh.de/colibri/about.html].
[3] DINUBE web page, access: [01.2016 https://www.dinube.com/en/what-is-dinube/].
[4] European Central Bank 2013. Recommendations for the security of mobile payments.
access: [11.01.2016 https://www.ecb.europa.eu/pub/pdf/other/recommendationssecurityinternetpaymentsoutcomeofpcfinalversionafterpc201301en.pdf].
[5] Goodin D., 2015. Severe weaknesses in Android handsets could leak user fingerprints, access: [09.01.2016 http://arstechnica.com/security/2015/08/severe-weaknesses-in-android-handsets-could-leak-user-fingerprints/].
[6] Klimontowicz M., 2014. Rynek płatności mobilnych w Polsce – stan i perspektywy rozwoju. Annales Universitatis Mariae Curie-Skłodowska vol. XLVIII (3) Lublin.

[7] Mikowska M., 2013. Marketing mobilny w Polsce. Jestem Mobi, Katowice.
[8] Mikowska M., 2015. Polska jest MOBI. Jestem Mobi, Katowice.
[9] Secutiy Research Labs. 2015. access: [09.01.2016 https://srlabs.de/spoofing-fingerprints/].
[10] Visa 2013. Mobile payment acceptance solutions. access: [11.01.2016 https://usa.visa.com/dam/VCOM/download/merchants/bulletin-mobile-best-practices.pdf].

POPRAWA ZABEZPIECZEŃ W SYSTEMIE PŁATNOŚCI MOBILNEJ

Streszczenie

W pracy opisano płatności mobilne i dostępne sposoby ich wykonywania za pomocą smartfonów. Artykuł zawiera analizę bezpieczeństwa płatności mobilnych, a także omawia potencjalne ryzyka kradzieży danych jakie są z nimi związane. Analiza przeprowadzona została na podstawie informacji dostępnych w Internecie oraz przeprowadzonych badań. W dokumencie zawarto również propozycję usprawnienia sposobu zabezpieczeń, która przy zachowaniu wygody mogłaby dobrze służyć użytkownikom płatności mobilnych. Artykuł jest oparty na badaniach przeprowadzonych w ramach projektu COLIBRI Erasmus+ oraz dostępne źródła internetowe.

Słowa kluczowe: system płatności mobilnej, bezpieczeństwo, COLIBRI Erasmus+

UNIWERSYTET TECHNOLOGICZNO-PRZYRODNICZY
IM. JANA I JĘDRZEJA ŚNIADECKICH W BYDGOSZCZY
ZESZYTY NAUKOWE NR 265
TELEKOMUNIKACJA I ELEKTRONIKA 19 (2016) 29-36

FACE RECOGNITION USING THE HAAR CLASSIFIER CASCADE AND FACE DETECTION BASED ON DETECTION OF SKIN COLOR AREAS

Gracjan Kątek, Agnieszka Holik, Tomasz Zabłocki, Pamela Dobrzyńska

UTP University of Science and Technology,
Faculty of Telecommunications, Computer Science and Electrical Engineering,
al. Prof. S. Kaliskiego 7, 85-796 Bydgoszcz, Poland
grakat001@utp.edu.pl, agnhol001@utp.edu.pl,
tomzab001@utp.edu.pl, pamdob000@utp.edu.pl

Summary. The article presents two methods of face detection. The first of these is a method Haar classifier cascade. The second is a face detection method based on detection of skin color areas. They propose a face detection algorithm based on skin color. The main emphasis lies on the effectiveness of the algorithm in order to properly recognize a human face. The results allowed to evaluate the effectiveness of the proposed method.

Keywords: face detection, Haar classifier cascade, face detection based on skin color

1. INTRODUCTION

Systems are used to classify recognition occurring in a real or artificial environment viewed by their models. Face recognition is based on specific functions performed by the algorithm [1, 2, 3]. By gradual processing of the image data, the algorithm tracks, analyzes and corrects the data [9].

An image that is delivered and on which the searched item exists (in this case the searched item is the face) contains difficult features, such as: lighting, which source is atypically located, colors, the object is not directed straight ahead, but in some other direction. In this case it is necessary to use a classifier [9].

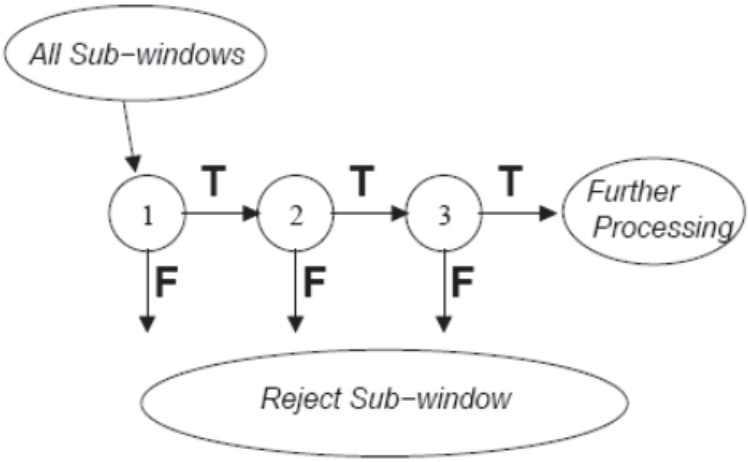

Fig. 1. The process of detection using a cascade classifier

Classification is the process of grouping objects in terms of similarities. Cascade classifier verifies by dividing this into stages (Figure 1). Haar classifier is to locate the searched object (here: the face) on the image where the face detection is performed by sliding the window along the image. A search box is moved along the picture to check whether the current area belongs to the object being sought. The algorithm is characterized by three features: edge (edge features), line (line features) and center surround (center-surround features). The main advantage of operating with Haar-like characteristics is their feature to compile information of the described areas under certain conditions. One example would be the edges or lines presented in a given area.

Face detection based on the detection of the image of skin color or skin-like color areas is an algorithm, which is to detect a field containing the characteristics of the face, such as skin color. Objects which meet certain conditions are filtered and are further analyzed and corrected. Shadows and different hue of light might cause difficulties. To meet the condition to find an area with characteristics similar to the human face including skin color, an appropriate color model of skin color must be adopted. In this case, the color palette for which the condition of detecting areas is fulfilled should be selected first. In that case the best is the TSL space, which is an area that is processed perceptually and determines the color as:

a) color, that is the degree to which a stimulus can be described as similar to or different from other stimuli described as red, green, yellow and white, and can be regarded as the tint with the addition of white [7]
b) saturation
c) light / brightness of a given stimulus, which appears to be white under similar viewing conditions or similar to it.

Both methods are effective when the face is directed vertically. In a situation where the face is pivoted the chance to recognize the decrease.

2. THE DETECTION ALGORITHM BASED ON SKIN COLOR

For test purposes the algorithm was developed to recognize the face based on skin color. For the purposes of testing algorithm was developed to recognize the face based on skin color.

1. Load the photo.
2. Convert photo to YCrCb and HSV formats.
3. Create skin mask which allows only 1 or 0 values, fill it with zeros.
4. Detect skin pixels based on below rules:
   a. for RGB format (both rules sets are combined by logical OR):
      – skin pixels in daylight are expressed by this rule:
        (R > 95 && G > 40 & B > 20 && MAX(R,G,B) - MIN (R,G,B) > 15 && |R-G| > 15 && R > G && R > B)
      – skin pixels in artificial light are expressed by this rule:
        (R > 220 && G > 210 && B > 170 && |R - G| <= 15 && R > B && G > B)
   b. for YCrCb format:
      – (Cr <= 1.5862 * Cb + 20 && Cr >= 0.3448 * Cb + 76.2069 && Cr > -4.5652 * Cb + 234.5652 && Cr <= -1.15 * Cb + 301.75 && Cr <= -2.2857 * Cb + 432.85)
   c. for HSV format:
      – (H < 17 && H > 162)
5. Mark each pixel, which fulfills above rules combined by logical AND, on our skin mask by value 1.
6. Do some morphological operations. In our algorythm we use Erode, Dilate and Closing.
7. Group pixels in blobs.
8. For each area greater than 500 detect if this area is face by above rules:
   a) (box_ratio >= 0.35 && box_ratio <= 1.1)
   b) (eccentricity >= 0.25 && eccentricity <= 0.95)
   c) (extent >= 0.35)

NOTE: Box Ratio is ratio of width to height of the bounding box, Eccentricity is the ratio of the minor axis to major axis of a bounding ellipsem and Extent is the ratio of area of region to area of bounding box.

9. For each region that fulfills above rules combined by logical AND draw bounding box around it (regions with drawed box are our detected faces).
10. Save the image.

3. THE TEST METHOD

The test method consist of testing both algorithms by running them on virtual machines. The machines were equipped with an installed library OpenCV [6, 8] and Python version 3.5. Ubuntu operating system was installed on each machine. The test

involved the face in the 25 photographs. To be sure the result of each test was repeated 10 times. Before performing the tests, a number of training classifiers with 5,000 positive and 5,000 negative samples has been completed. An important element of preparation, since the task was to train a classifier. The training set consisted of 100 positive and 100 negative samples. The samples had a size of 50x50 pixels. In the picture (Fig. 2) was presented a positive sample in gray scale and Fig. 3 presented a negative sample. Examples of test samples are shown in the figures (Figs. 4-6).

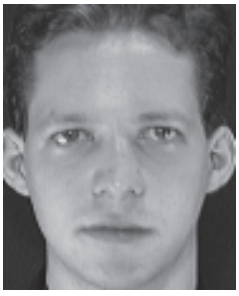

Fig. 2. Sample of positive

Fig. 3. Sample of negative

Fig. 4. A group of people on an inhomogeneous background

Fig. 5. A group of people on an homogeneous background

Fig. 6. Person on an homogeneous background

As shown in Figures 4-6, tested different types of images - because of the background color and the number of people in the pictures. We have not been studied images in which the faces are not compatible with a vertical axis of the picture.

4. RESULTS

Table 1 shows the results of the simulation for face detection based on skin color. As may be seen, this method works properly when in the picture is one person. If there are more people, algorithm effectiveness decreases. Based on these results it can be concluded that the effectiveness of the algorithm is of 90%.

Table 1. The results of detection based on skin color

Figure Number	**Faces on figure**	**Total number**	**No. of false**	**No. of positive**	**False detection**	**Detection summary**
1	1	1	0	1	0%	100%
2	1	1	1	0	100%	0%
3	1	1	0	1	0%	100%
4	1	1	0	1	0%	100%
5	1	1	0	1	0%	100%
6	1	0	0	0	0%	0%
7	2	4	2	3	50%	100%
8	4	4	0	2	0%	100%
9	1	3	2	1	67%	100%
10	8	5	1	5	20%	63%
11	12	14	3	11	21%	92%
12	8	8	0	8	0%	100%
13	3	3	1	2	33%	67%
14	1	3	0	2	0%	200%
15	1	2	0	1	0%	100%
16	2	4	2	2	67%	100%
17	2	3	1	2	50%	100%
18	1	1	2	1	50%	100%
19	1	1	0	1	33%	100%
20	1	0	0	1	0%	100%
21	1	1	0	0	0%	0%
22	1	1	0	1	0%	100%
23	1	1	0	1	0%	100%
24	4	4	0	4	0%	100%
25	1	1	0	1	0%	100%
				Average	**14%**	**90%**

These results are shown in figure 7. It can be seen that errors in the recognition of faces appeared for images from 8 to 18. As mentioned earlier, these images were more than one person.

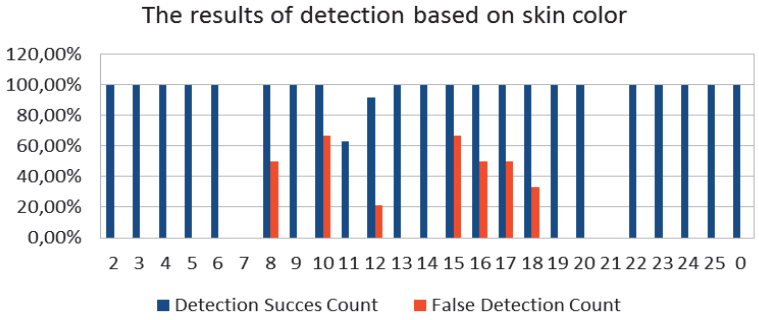

Fig. 7. Detection based on skin color

Table 2 shows the results of Haar classifiers algorithm. Tests were performed on the same set of images. For this algorithm, it is more important the orientation of face item relative to the vertical axis than the number of persons in the photo. The effectiveness of this algorithm is 92%.

Table 2. Haar classifier cascade

Figure Number	**Faces on figure**	**Total number**	**No. of false**	**No. of positive**	**False detection**	**Detection summary**
1	1	1	0	1	0%	100%
2	1	1	1	0	100%	0%
3	1	1	0	1	0%	100%
4	1	1	0	1	0%	100%
5	1	1	0	1	0%	100%
6	1	0	0	0	0%	0%
7	2	4	1	3	25%	150%
8	4	4	0	4	0%	100%
9	1	3	2	1	67%	100%
10	8	5	1	4	20%	50%
11	12	14	3	11	21%	92%
12	8	8	0	8	0%	100%
13	3	3	1	2	33%	67%
14	1	3	1	2	33%	200%
15	1	2	1	1	50%	100%

Table 2 cont.

16	2	4	1	3	25%	150%
17	2	3	1	2	33%	100%
18	1	1	2	1	200%	100%
19	1	1	0	1	0%	100%
20	1	0	0	1	0%	100%
21	1	1	1	0	100%	0%
22	1	1	0	1	0%	100%
23	1	1	0	1	0%	100%
24	4	4	0	4	0%	100%
25	1	1	0	1	0%	100%
Average					**28%**	**92%**

Results of the simulation are shown in figure 8. As shown, this algorithm exhibits a much more errors than the detection discussed previously. Its higher effectiveness due to the better facial recognition on photos with more people.

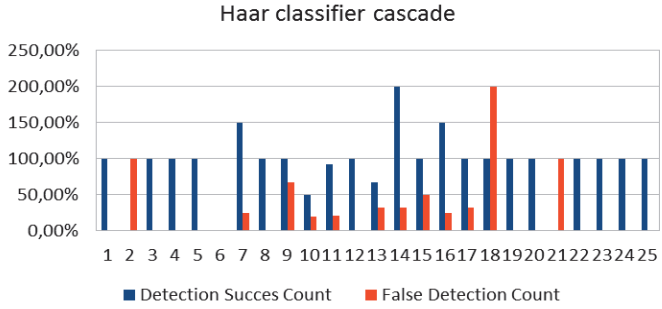

Fig. 8. Detection based on Haar classifier cascade

5. CONCLUSION

Based on the results in tables (Tab. 1. Tab. 2) and figure (Fig. 7. Fig. 8) it can be concluded that both methods are equally effective for color images. Although in the study the method based on Haar classifier cascade showed a higher percentage of errors, it can be more effective because it is able to recognize the face in black and white. Thus, it is more flexible than the other methods of detection.

BIBLIOGRAPHY

[1] http://www.cs.put.poznan.pl/kkrawiec/piro-projects/2006-1/ro.pdf- Rozpoznawanie obrazów: Wykrywanie orientacji zdjęć przez lokalizację twarzy s. 2-5
[2] http://flash.iiar.pwr.edu.pl/~jkedzier/download/archiwum/2012/ matkowski_sobczak.pdf.
[3] http://pl.wikipedia.org/wiki/Rozpoznawanie_wzorc%C3%B3w
[4] http://rab.ict.pwr.wroc.pl/~mw/Stud/Dypl/lkucharczyk/wykrywa nie_twarzy_praca_dyplomowa _2011.pdf 23-25.
[5] http://pl.wikipedia.org/wiki/YCbCr
[6] http://docs.opencv.org/_images/haarfeatures.png
[7] http://en.wikipedia.org/wiki/TSL_color_space
[8] http://en.wikipedia.org/wiki/Hue
[9] Choraś R.S., 2005. Komputerowa wizja. Metody interpretacji i identyfikacji obiektów, Wydawnictwo Exit.

ROZPOZNAWANIE TWARZY METODĄ KASKADY KLASYFIKATORÓW HAARA I DETEKCJA TWARZY W OPARCIU O WYKRYWANIE OBSZARÓW O KOLORZE SKÓRY

Streszczenie

W artykule przedstawiono dwie metody detekcji twarzy. Pierwsza z nich to metoda kaskady klasyfikatorów Haara. W metodzie tej ważne jest położenie twarzy w stosunku do kąta obrócenia zdjęcia. Rozpoznawane są tylko „pionowe" twarze. Drugą stanowi metoda detekcji twarzy w oparciu o wykrywanie obszarów o kolorze skóry. Zaproponowano algorytm detekcji twarzy w oparciu o kolor skóry. Główny nacisk położono na skuteczność algorytmu w celu poprawnego rozpoznania ludzkiej twarzy. Otrzymane wyniki pozwoliły ocenić skuteczność zaproponowanej metody.

Keywords: rozpoznawanie twarzy, kaskady klasyfikatorów Haara, detekcja twarzy na podstawie koloru skóry