JOURNAL of COMPUTER SCIENCES INSTITUTE

JCSI 8 (2018) 215-219
Wysłane: 2017-12-14
Przyjęte: 2017-12-19

Rozpoznawanie obiektów w obrazie wideo z kamery do komputera

Oleksandr Cherednyk*, Elżbieta Miłosz

Politechnika Lubelska, Instytut Informatyki, Nadbystrzycka 36B, 20-618 Lublin, Polska

Streszczenie: Celem artykułu jest określenie skuteczności wykrywania obiektu w obrazie wideo za pomocą kamery internetowej. W trakcie pracy zostały zbadane i opisane podstawowe metody rozpoznawania obiektów na zdjęciu, a mianowicie wykorzystanie sztucznych sieci neuronowych i metod Viola-Jones. Do przeprowadzenia eksperymentów, na podstawie metody Viola-Jones, realizowano aplikację do rozpoznawania obiektów na wideo. Za pomocą tej aplikacji, przeprowadzono badanie w celu określenia skuteczności metody Viola-Jones do wykrywania obiektów na wideo.

Słowa kluczowe: Viola-Jones; gest; rozpoznawanie

*Autor do korespondencji.

Adres e-mail: oleksandr.cherendyk@gmail.com

Object recognition on video from camera to computer

Oleksandr Cherednyk*, Elżbieta Miłosz

Institute of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract. The goal is to determine the effectiveness of object detection in a video using the camera for the computer. In the course of work studied and described the main methods of recognition of objects in the image, namely the use of artificial neural networks and techniques of Viola-Jones. For the study, based on the method of Viola-Jones, implemented the application for object recognition in video, as this method is effective for solving this problem. With this application, a study was conducted to determine the effectiveness of the method of viola-Jones to detect objects in the video.

Keywords: Viola-Jones; gesture; recognition

*Corresponding author.

E-mail address: oleksandr.cherendyk@gmail.com

1. Wprowadzenie

Ludzie nieustannie borykają się z problemem rozpoznawania obrazów. Ludzki mózg radzi sobie z tym zadaniem bez problemu. Ale istnieją sytuacje, kiedy człowiek potrzebuje pomocy komputera, aby rozpoznawać obiekty. W dzisiejszych czasach komputer jest często używany do rozpoznawania obiektów. Technologia ta jest na przykład wykorzystywana w ruchu drogowym, w celu wyszukiwania numerów rejestracyjnych samochodów, na lotniskach dla wyszukiwania twarzy przestępców, a także do rozpoznawania gestów, pozwalających na realizację interakcji człowieka z komputerem. Pomimo tego, że technologia i narzędzia rozpoznawania obrazów osiągnęły dobry poziom, to nie są one doskonałe i wymagają dalszych badań.

2. Metody rozpoznawania obrazów

Wideo jest to sekwencja obrazów. Dlatego, aby rozpoznać obiekt na wideo, należy rozpoznać go w jednym obrazie. Jednymi z głównych sposobów rozpoznawania obiektu na obrazie są:

- zastosowanie metody Viola-Jones;
- wykorzystanie sztucznych sieci neuronowych.

W 2001 roku Paul Viola i Michael Jones zaproponowali algorytm pozwalający wykrywać obiekty na obrazie w czasie rzeczywistym (metoda Viola-Jones) [2][3]. Metoda pozwala wykrywać różne typy obiektów. Istnieją gotowe realizacje i ulepszenia tej metody, jedna z nich jest dostępna w zawartości biblioteki OpenCV [1]. Zalety metody Viola-Jonsa to przede wszystkim wysoka dokładność rozpoznawania obrazu i możliwość wyszukiwania obiektów na obrazie w czasie rzeczywistym. W warunkach pracy pod niewielkim kątem (do 30 stopni), metoda Viola-Jones ma wysoką szybkość rozpoznawania, a także skutecznie działa w różnych warunkach oświetlenia [4].

Praca algorytmu metody Viola-Jones opiera się na czterech koncepcjach [5]:

1. Reprezentacja obrazu w integralnej postaci, co pozwala szybko obliczyć niezbędne cechy obiektów.
2. Wyszukiwanie żądanego obiektu na podstawie prostokątnych kształtów zwanych falkami (cechami) Haara.
3. Zastosowanie algorytmu "boosting", co pozwala wybrać bardziej odpowiednie cechy obiektu w określonej części obrazu.
4. Korzystanie z kaskadowego filtrowania okien, gdzie nie znaleziono obiektu.

Zintegrowany obraz jest przedstawiony w postaci dwuwymiarowej macierzy, której rozmiar jest równy rozmiarowi przychodzącego obrazu. Obraz cyfrowy przechowuje w sobie wartość koloru piksela, dla obrazu czarno-białego, wartość od 0 do 255, a dla kolorowego wynosi od 0 do ⟦255⟧^3(wartości R, G, B). W integralnej postaci każdy element macierzy przechowuje w sobie sumę natężenia wszystkich pikseli znajdujących się na lewo

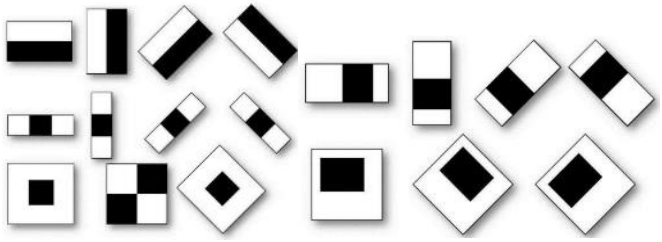
i powyżej tego elementu. Obliczenie elementów macierzy odbywa się za pomocą następującego wzoru [3]:

$$L(x,y) = \sum_{i=0,j=0}^{i\le x,j\le y} *I(i,j), \quad (1)$$

gdzie I(i, j) – jasność piksela obrazu źródłowego. Każdy element integralnego obrazu L(x, y) zawiera w sobie sumę wszystkich pikseli w prostokątnym odcinku od (0, 0) do (x, y) [4]. Obliczanie integralnego obrazu możliwe jest za pomocą następującego wzoru [10]:

$$L(x,y) = I(x,y) - L(x-1,y-1) + L(x,y-1) + L(x-1,y), \quad (2)$$

Cechy Haara, są używane do wyszukiwania obiektu. Obraz jest przetwarzany częściami, okno o określonej wielkości porusza się na obrazie i dla każdego odcinka przebytego obrazu, określa klasyfikatory Haara. Dostępność obiektu na obrazie zależy od różnicy między wartością cechy. W większości są używane prostokątne znaki (Rysunek 1).



Rys 1. Cechy wykorzystywane w klasyfikatorach Haara, po lewej główne oznaki, po prawej dodatkowe oznaki [9]

Wartość funkcji jest obliczana za pomocą wzoru:

$$F = X - Y, \quad (3)$$

gdzie: X – suma wartości jasności punktów na odcinku zamkniętym jasną częścią znaku, Y – suma wartości jasności punktów na odcinku zamkniętym ciemnym częścią znaku.

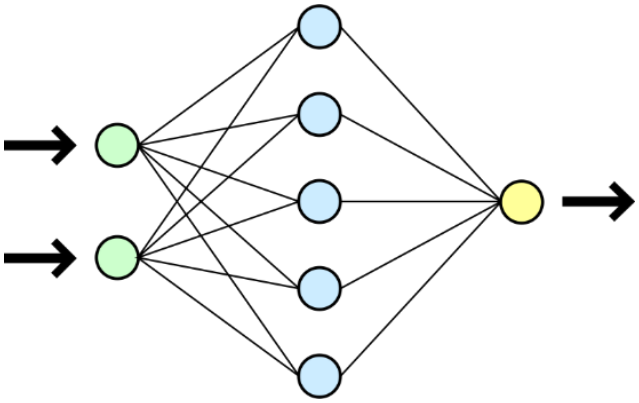
Do wyboru najbardziej odpowiednich cech szukanego obiektu na części obrazu wykorzystywany jest algorytm "boosting". "Boosting" oznacza poprawę. Metoda ta zwiększa skuteczność algorytmu uczenia, pozwala zbudować silny złożony klasyfikator łącząc słabe i proste klasyfikatory.

Bardziej ulepszony algorytm "boosting" – metoda AdaBoost, została zaproponowana w 1999 roku przez Yoavema Freundoem i Roberta Schapirem. Strumień wideo, uzyskany za pomocą kamery, jest sekwencją klatek. Dla każdego kadru jest obliczany jego zintegrowany obraz. Następnie kadr skanuje się oknem małych rozmiarów, zawierających znaki Haar. Dla każdego i-tego znaku odpowiedni kwalifikator określa wzór [4]:

$$h(z) = \begin{cases} 1, p_i f_i(z) < p_i Q_i \\ 0 \end{cases}, \quad (4)$$

gdzie: z – okno, Q_i – wartość progowa, p_i – kierunek znaku nierówności, f_i – cecha Haara.

Sztuczna sieć neuronowa to połączenie równoległo-szeregowe neuronów za pomocą synaps. Jej model matematyczny, realizowany za pomocą oprogramowania, jest mocno uproszczonym model mózgu (Rysunek 2).



Rys 2. Przykład sztucznej sieci neuronowej [6]

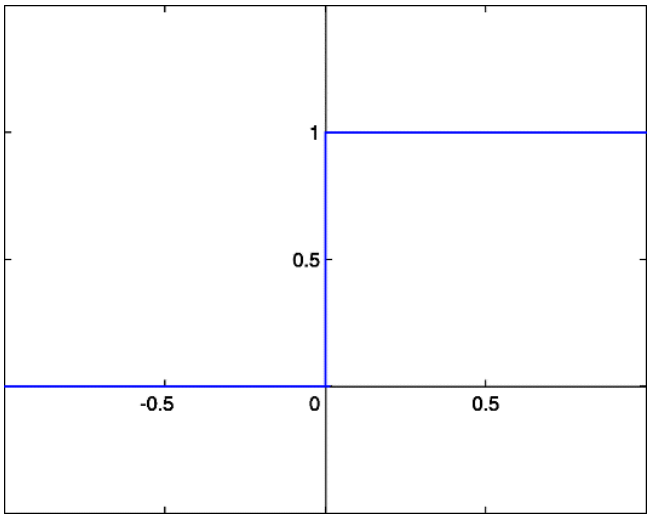
Sztuczne neuronowe sieci stosowane są do rozwiązywania trudnych zadań, wymagających analitycznego obliczenia, takich jak: klasyfikacja, przewidywanie i rozpoznawanie. Rozpoznawanie jet chyba najbardziej szerokim zastosowaniem sztucznych sieci neuronowych.

Przykładowa sieć neuronowa (Rysunek 2) składa się z neuronów i synaps. Posiadaona trzy warstwy neuronów: wejściową, ukrytą i wyjściową. W wejściowej warstwie podawane są dane wejściowe, a następnie w ukrytej warstwie odbywa się rozpoznawanie, po czym wyniki są przesyłane do warstwy wynikowej.

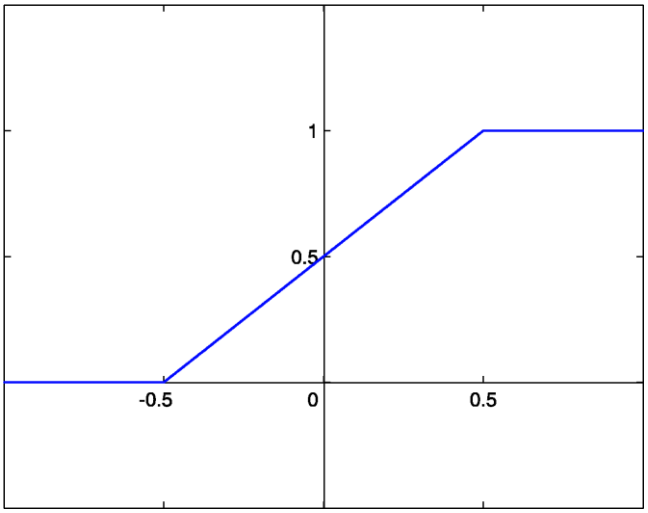
Neuron reprezentuje jednostkę obliczeniową, która wyznacza ważona sumę wartości wejściowych. Synapsa to łącznik neuronów, który ma swoją wagę. Sygnał wyjściowy zależy od sumy sygnałów wejściowych, pomnożonych przez ich wagi. W zależności od funkcji aktywacji, otrzymaną wartość aktywuje neuron i na wyjście będzie podawany sygnał 1, albo neuron nie zostanie aktywowany i w tym przypadku na wyjście będzie podawany sygnał 0.

Podstawowe funkcje aktywacji (Rysunki 3,4,5) to:

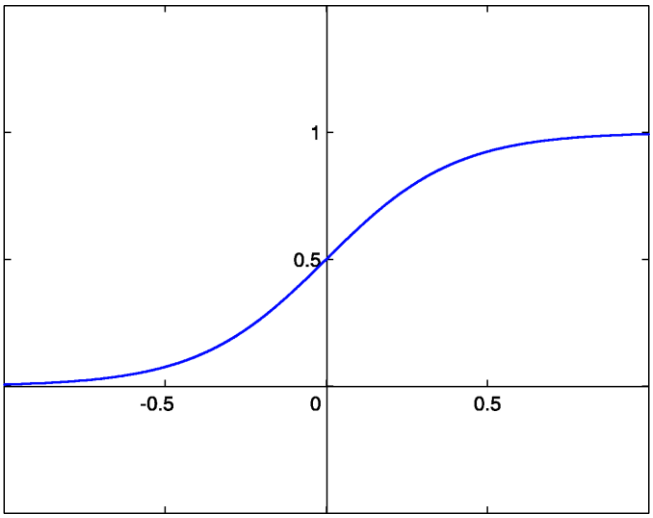
- graniczna funkcja,
- funkcja liniowa,
- "Sigmoidalna" funkcja.



Rys 3. Graniczna funkcja aktywacji [7]

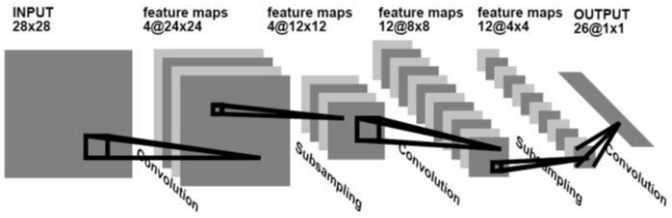


Rys 4. Liniowa funkcja aktywacji [7]



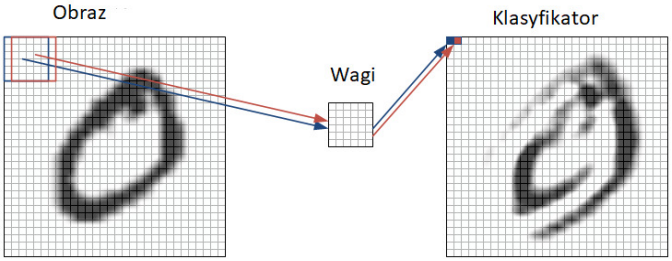
Rys 5. "Sigmoidalna" funkcja aktywacji [7]

Do pracy z obrazem można używać splotowe sieci neuronowe. Splotowe sieci neuronowe – to jeden z rodzajów sieci neuronowych często używany w komputerowym przetwarzaniu obrazów (Rysunek 6.) [11].



Rys 6. Praca splotowych sieci neuronowych [8]

Obraz uzyskany z kamery wchodzi w sieć neuronową nie w pełni, a rozbija się na obszarze o wymiarach n*n. Następnie te części po kolei przekazywane do pierwszej warstwy sieci neuronowej. Wszystkie przychodzące obszary zdjęcia są mnożone przez małą macierz wag. Rozmiar tej macierzy odpowiada przychodzącemu obszarowi obrazu, czyli n*n, taka macierz nazywa się jądrem. Wyniki mnożenia są sumowane i zapisywane w podobne stanowisko przychodzącego obrazu (Rysunek 7.).



Rys 7. Przykład zestawienia obrazu [8]

Każdy fragment obrazu element po elemencie jest mnożony przez niewielką macierz wag (rdzeń), wyniki są sumowane. Wynikiem mnożenia zostanie odebrana karta cech [11].

3. Cel i plan badań

Celem badań jest określenie skuteczności wykrywania obiektu na wideo za pomocą kamery internetowej podłączonej do komputera. Należy potwierdzić lub odrzucić następującą hipotezę:

Biblioteka OpenCV wykorzystującą metodę Viola-Jones pozwala na realizację komputerowego interfejsu sterowania gestami dłoni lub obrazem twarzy poprzez skuteczne rozpoznawanie obrazów dłoni lub twarzy na wideo za pomocą kamery podłączonej do komputera.

Do przeprowadzenia badań, została napisana aplikacja (Listing 1) z wykorzystaniem biblioteki OpenCV, która pozwala rozpoznawać obiekty za pomocą kamery podłączonej do komputera.

Listing 1. Kod programu

```
using System;
using System.Drawing;
using System.Windows.Forms;
using Emgu.CV;
using Emgu.CV.Structure;
using Emgu.CV.CvEnum;
namespace proga {
public partial class Form1 : Form {
private Capture capture;
private HaarCascade haarCascade;
int centr = 1;
public Form1() {
InitializeComponent();
}
private void timerTick(object sender, EventArgs e) {
using (Image<Bgr, byte> image = capture.QueryFrame()) {
if (image != null){
Image<Gray, byte> grayImage = image.Convert<Gray,
byte>();
var hands = grayImage.DetectHaarCascade(haarCascade,
1.4, 4, HAAR_DETECTION_TYPE.DO_CANNY_PRUNING,
new Size(image.Width / 8, image.Height / 8))[0];
foreach (var hand in hands){
image.Draw(hand.rect, new Bgr(0, double.MaxValue, 0), 3);
}
 pictureBox1.Image = image.ToBitmap();
}}}
private void btnStart_Click(object sender, EventArgs e){
capture = new Capture(0);
haarCascade = new HaarCascade("D:\\
haarcascade_frontalface_alt.xml.xml");
Application.Idle += timerTick;
}
private void label1_Click(object sender, EventArgs e){}
```

```
private void pictureBox2_Click(object sender, EventArgs e){}
}}
```

W celu określenia skuteczności rozpoznawania gestów przeprowadzono szereg eksperymentów. Zostały one przeprowadzone z udziałem 8 osób. Jako urządzenie został użyty komputer ze wbudowaną kamerą o rozdzielczości 1280x720 pikseli.Do pierwszego eksperymentu jest wykorzystywany klasyfikator do rąk, wyszkolony na przykładzie 500 pozytywnych obrazów rąk i 1000 negatywnych obrazów, na których nie ma rąk (tło i inny obiekty). Do drugiego i trzeciego eksperymentu użyto klasyfikator do rąk z biblioteki OpenCV. Do czwartego eksperymentu użyto klasyfikator do twarzy z biblioteki OpenCV. Badania przeprowadzono w pomieszczeniach z obecnością innych przedmiotów z uwzględnieniem różnego oświetlenia i różnej odległości od użytkownika do komputera. W trakcie badań każdy uczestnik pokazywał polecenie 10 razy. Wykonano 10 pomiarów, z których 7 pochodzi od 7 uczestników, i 3 pomiary przeprowadzone zostały przez jednego uczestnika. Wynikami badania są 3 wartości:

- a - rozpoznawanie;
- b - nie rozpoznawanie;
- c - mylne rozpoznawanie;

4. Wyniki badań

Do pierwszego eksperymentu użyto klasyfikator wyszkolonego za pomocą 500 pozytywnych zdjęć rąk i 1000 negatywnych zdjęć. Eksperyment jednak nie powiódł się, ponieważ klasyfikator był słabo nauczony.

Drugi eksperyment rozpoznawania ręki przeprowadzono w pomieszczeniu z naturalnym oświetleniem, podczas słonecznego dnia. Użyto klasyfikatora z biblioteki OpenCV. Jego wyniki przedstawiono w tabeli 1.

Tabela 1. Wyniki eksperymentu wykrywania ręki w słoneczną pogodę

Numer badania	Liczba wyników a/b/c				
	50cm	1m	1m 50cm	2m	2m 50cm
1	10/0/0	9/1/0	7/3/0	4/6/0	-
2	9/1/0	10/0/0	9/1/0	3/7/1	-
3	10/0/0	9/1/0	9/1/1	5/5/0	-
4	10/0/0	10/0/0	8/2/0	3/7/0	-
5	10/0/1	10/0/1	10/0/0	2/8/1	-
6	9/1/0	8/2/1	8/2/0	3/7/0	-
7	10/0/0	10/0/0	6/4/0	4/6/0	-
8	10/0/0	10/0/0	9/1/0	3/7/0	-
9	9/1/1	9/1/0	7/3/1	5/5/1	-
10	10/0/0	9/1/0	8/2/0	3/7/0	-

Z analizy wyników (Tabela 1) widać, że skuteczne rozpoznawanie ręki odbywa się w odległości do 1-go metra. Program działa normalnie w odległości do 1m 50cm. W odległości 2 metrów program działa źle, ponieważ współczynnik rozpoznawania jest niski. Praca z komputerem na odległość ponad 2 metrów nie jest możliwa, ponieważ rozpoznawanie ręki praktycznie nie zachodzi.

Trzeci eksperyment rozpoznawania ręki przeprowadzono w pomieszczeniu z naturalnym oświetleniem, w czasie wieczornej pory dnia. Użyto klasyfikatora z biblioteki OpenCV. Wyniki badań przedstawiono w tabeli 2.

Tabela 2. Wyniki eksperymentu wykrywania ręki w godzinach wieczornych

Numer badania	Wyniki a/b/c			
	50sm	1m	1m 50sm	2m
1	9/1/0	10/0/0	8/2/0	-
2	10/0/0	9/1/0	8/2/0	-
3	10/0/0	8/2/0	9/1/0	-
4	9/1/1	9/1/1	7/3/0	-
5	8/2/0	9/1/0	8/2/0	-
6	9/1/0	10/0/0	6/4/0	-
7	10/0/0	8/2/0	8/2/0	-
8	9/1/0	8/2/0	9/1/0	-
9	10/0/0	9/1/0	8/2/0	-
10	9/1/0	9/1/0	7/3/0	-

Prawidłowe rozpoznawanie ręki zachodzi w odległości do 1m 50cm. Ustalono, że słabsze oświetlenie zmniejszyło liczbę fałszywych rozpoznawań, ale jednocześnie zmniejszyła się skuteczność rozpoznawania ręki.

Następny eksperyment przeprowadzono przez rozpoznawanie twarzy. Był używany klasyfikator z biblioteki OpenCV. Rozmiar pliku klasyfikatora dla twarzy był kilka razy większy niż rozmiar pliku klasyfikatora dla rąk. Oznacza to, że dla nauczenia klasyfikatora twarzy użyto znacznie większą liczbę pozytywnych i negatywnych próbek obiektu. Wyniki badań przedstawiono w tabeli 3.

Tabela 3. Wyniki eksperymentu wykrywania twarzy

Numer badania	Wyniki a/b/c				
	50cm	1m	1m 50cm	2m	2m 50cm
1	10/0/0	10/0/0	10/0/0	10/0/0	-
2	10/0/0	10/0/0	10/0/0	10/0/0	-
3	10/0/0	10/0/0	10/0/0	10/0/0	-
4	10/0/0	10/0/0	10/0/0	9/1/0	-
5	10/0/0	10/0/0	10/0/0	10/0/0	-
6	10/0/0	10/0/0	10/0/0	10/0/0	-
7	10/0/0	10/0/0	10/0/0	10/0/0	-
8	10/0/0	10/0/0	10/0/0	9/1/0	-
9	10/0/0	10/0/0	10/0/0	10/0/0	-
10	10/0/0	10/0/0	10/0/0	10/0/0	-

Wyniki eksperymentu dotyczące wykrywania twarzy, okazały się najlepsze. Rozpoznawanie odbywa się na podobnej odległości, jak i rozpoznawania ręki. Ale przy tym rozpoznawanie twarzy ma wysoką skuteczność. Przy tym nie było żadnych fałszywych wykryć.

5. Wnioski

Do przeprowadzenia skuteczności rozpoznawania obiektów za pomocą kamery internetowej, została wybrana biblioteka OpenCV wykorzystująca metodę Viola-Jones, tak jak przypuszczano biblioteka była skuteczna dla rozwiązania tego zadania. W trakcie badań wykorzystano trzy karty cech i dwa rodzaje obiektów – ręka i twarz. W trakcie pracy z mniej

wyuczonym klasyfikatorem, wyniki wykazały niski poziom rozpoznawania. Podczas korzystania z lepiej wyuczonego klasyfikatora, wyniki wykazały wysoki poziom rozpoznawania. W trakcie badań z wykorzystaniem dobrze wyuczonego klasyfikatora do rozpoznawania twarzy, wyniki wykazały bardzo wysoką skuteczność. Wyniki badań potwierdziły założoną hipotezę: Biblioteka OpenCV wykorzystująca metodę Viola-Jones pozwala na realizację komputerowego interfejsu sterowania gestami dłoni lub obrazem twarzy poprzez skuteczne rozpoznawanie obrazów dłoni lub twarzy na wideo za pomocą kamery podłączonej do komputera. Jakość rozpoznawania zależy od jakości wyuczenia klasyfikatora:im lepiej wyuczony klasyfikator, tym wyższa skuteczność rozpoznawania.

Literatura

[1] Monali Chaudhari, Shanta Sondur, Gauresh Vanjare. «A review on Face Detection and study of Viola Jones method». International Journal of Computer Trends and Technology (IJCTT), 2015.

[2] Paul Viola , Michael Jones. “Rapid object detection using a boosted cascade of simple features”. Accepted conference on computer vision and pattern recognition. 2001.

[3] Paul Viola , Michael Jones. “Robust Real-time Object Detection”. International Journal of Computer Vision. 2004.

[4] Мурлин А.Г., Пиотровский Д.Л., Руденко Е.А., Янаева М.В. «Алгоритмы и методы обнаружения и распознавания жестов руки на видео в режиме реального времени». Научный журнал КубГАУ, 2014.

[5] Спицын В.Г., Буй Тхи Тху Чанг, Фан Нгок Хоанг. «Распознавание лиц на основе применения метода Виолы-Джонса, вейвлет-преобразования и метода главных компонент». Томский политехнический университет, 2012.

[6] https://commons.wikimedia.org/wiki/File:Neural_network.svg [13.12.2017]

[7] https://dic.academic.ru/dic.nsf/ruwiki/19743 [13.12.2017]

[8] https://geektimes.ru/post/74326/ [13.12.2017]

[9] https://habrahabr.ru/post/133826/ [13.12.2017]

[10] Метод Виолы Джонса как основа для распознавания лиц https://habrahabr.ru/post/133826/ [13.12.2017]

[11] Применение нейросетей в распознавании изображений. https://geektimes.ru/post/74326/ [13.12.2017]