



Politechnika
Śląska

POLITECHNIKA ŚLĄSKA
WYDZIAŁ AUTOMATYKI, ELEKTRONIKI I INFORMATYKI
KIERUNEK: AUTOMATYKA I ROBOTYKA

Praca dyplomowa inżynierska

Tytuł pracy dyplomowej inżynierskiej

autor: Szymon Ciemala

kierujący pracą: dr inż. Krzysztof Jaskot

konsultant: dr inż. Imię Nazwisko

Gliwice, grudzień 2021

Spis treści

1	Streszczenie	1
	Streszczenie	1
2	Wstęp	3
2.1	Wprowadzenie w problem	4
2.2	Cel pracy	4
2.3	Osadzenie problemu w dziedzienie	4
2.4	Charakterystyka rozdziałów	4
3	[Analiza tematu]	7
3.1	Sformułowanie problemu	7
3.2	Podobne rozwiązania	8
4	Wymagania i narzędzia	9
4.1	Wymagania	9
4.2	Narzędzia	10
4.2.1	System operacyjny Ubuntu	10
4.2.2	Python	10
4.2.3	iPython	10
4.2.4	Jupyter Notebook	10
4.2.5	Visual Studio Code - Środowisko Pracy	11
4.2.6	Platforma PyPi	11
4.2.7	Programy bdist_wheel oraz sdist	11
4.3	Metodyka Pracy	11
4.3.1	Praca z Jupyter Notebook	11
5	[Właściwy dla kierunku - np. Specyfikacja zewnętrzna]	13
5.1	Wymagania sprzętowe	13
5.1.1	Kamera	13
5.2	Wymagania programowe	13
5.3	Instalacja paczki	14

5.4	Przykładowy program	15
5.5	Funkcjonalność modułu	17
5.5.1	Parametry	17
5.5.2	Dostępne funkcje	18
5.5.3	Główny wątek	18
5.6	Przykłady użycia	18
5.6.1	Rozpoznawanie alfabetu w języku migowym	18
5.6.2	Interaktywny Kiosk	20
5.6.3	Dobór koloru	22
5.7	Stworzenie nowego modelu przez użytkownika	24
6	[Właściwy dla kierunku - np.Specyfikacja wewnętrzna]	25
6.1	Budowa klasy	25
6.2	Struktury danych	26
6.2.1	Dataclass	26
6.2.2	Słownik	26
6.2.3	Pickle	27
6.3	Rozpoznawanie dłoni	27
6.3.1	OpenCV - przygotowanie obrazu z kamery	27
6.3.2	MediaPipe - Elementy charakterystyczne	28
6.3.3	Generowanie grafiki dłoni	29
6.4	Pomiary oraz inne ważne elementy	29
6.4.1	Pozycja dłoni	29
6.4.2	Obrót dłoni	29
6.4.3	Odległość między palcami	29
6.5	Rozpoznawanie gestów	29
6.5.1	Przygotowanie modeli uczenia maszynowego	30
6.5.2	Zebranie danych	31
6.5.3	Budowa pliku CSV	34
6.5.4	Metody klasyfikacji - uczenie maszynowe	34
6.5.5	Wybrane algorytmy klasyfikujące	35
6.5.6	Badanie dokładności każdego z algorytmów	36
6.5.7	Ponowne wykorzystanie modelu	36

6.6	Paczka PyPi	36
6.6.1	Budowa paczki	36
6.6.2	Struktura Paczki	36
6.6.3	Pliki Konfiguracyjne	37
6.6.4	Załadowanie paczki do repozytorium	37
7	Weryfikacja i walidacja	39
8	Podsumowanie i wnioski	41
	Bibliografia	41
	Spis skrótów i symboli	45
	Źródła	47
	Zawartość dołączonej płyty	49

Rozdział 1

Streszczenie

Praca inżynierska „Rozpoznawanie obiektów z wykorzystaniem biblioteki OpenCV”, łączy tematykę wizji komputerowej oraz algorytmów uczenia maszynowego.

Praca skupia się na stworzeniu wygodnej w użyciu oraz powszechnie dostępnej biblioteki umożliwiającej wykorzystanie gestów, pozycji dłoni, obrotu dłoni oraz odległości między palcem wskazującym a kciukiem jak elementów sterujących w dowolnych projektach napisanych w języku Python.

Do napisania pracy wykorzystano biblioteki języka Python o otwartym kodzie źródłowym, głównie OpenCV, MediaPipe oraz SciKit Learn.

Rozdział 2

Wstęp

- wprowadzenie w problem/zagadnienie
- osadzenie problemu w dziedzinie
- cel pracy
- zakres pracy
- zwięzła charakterystyka rozdziałów
- jednoznaczne określenie wkładu autora, w przypadku prac wieloosobowych
 - tabela z autorstwem poszczególnych elementów pracy

2.1 Wprowadzenie w problem

Rozwój technologii w ostatnich czasach przyczynił się do coraz częstszego wykorzystywania wizji komputerowej oraz metod uczenia maszynowego do rozpoznawania oraz klasyfikacji różnego typu obiektów, w tym części ludzkiego ciała. Pozwala to na interakcję człowieka z aplikacjami, często w sposób bardziej naturalny.

2.2 Cel pracy

Projekt inżynierski ma na celu stworzenie biblioteki w języku Python, która pozwoli na przystępne wykorzystanie algorytmów rozpoznawania gestów oraz ruchu dłoni. Biblioteka powinna oferować gotowe rozwiązania, na przykład przygotowane modele matematyczne pozwalające na rozpoznawanie języka migowego oraz gestów podstawowych. Dodatkowo powinna pozwolić na wyznaczenie pozycji dłoni, jej typu oraz wartości charakterystycznych, na przykład odległości między końcówkami wybranych palców czy kąta obrotu dłoni.

2.3 Osadzenie problemu w dziedzinie

Aktualnie istnieją częściowo gotowe rozwiązania pozwalające na rozpoznanie i klasyfikację dłoni - biblioteka MediaPipe.

2.4 Charakterystyka rozdziałów

W pierwszym rozdziale zostanie poruszony temat genezy problemu oraz jego sformułowania. Zostaną dodatkowo opisane jego założenia wraz z wymaganą funkcjonalnością projektu.

W drugim temacie zostaną opisane techniczne tworzonej biblioteki oraz narzędzia wymagane do jej stworzenia. Zakres opisywanych narzędzi rozpocznie się od opisu wybranego systemu operacyjnego do programów służących do stworzenia pobieralnej paczki na platformi PyPi.

W rozdziale nr 5 zostaną opisane wymagania sprzętowe wymagane do poprawnego działania funkcji biblioteki. Przedstawione zostanie działanie klasy wraz

z jej parametrami oraz sposób jego wykorzystania. W drugiej jego części zostaną przedstawione przykłady wykorzystania modułu. Przykładowym projektami będą: program rozpoznający gest, interaktywny kiosk bezdotykowy i system doboru koloru przy pomocy gestu.

Kolejny rozdział opisuje budowę klasy wraz z najważniejszymi algorytmami oraz strukturami danych. Dodatkowo opisany zostanie Jupyter Notebook służący do generowania modeli uczenia maszynowego rozpoznających gest dłoni.

Rozdział 3

[Analiza tematu]

- sformułowanie problemu
- osadzenie tematu w kontekście aktualnego stanu wiedzy (state of the art) o poruszonym problemie
- studia literaturowe [?, ?, ?, ?] - opis znanych rozwiązań (także opisanych naukowo, jeżeli problem jest poruszany w publikacjach naukowych), algorytmów,

3.1 Sformułowanie problemu

Problem można podzielić na trzy części. Każda z nich odpowiada za wybraną część funkcjonalności biblioteki.

- Wyznaczenie pozycji dłoni, odległości pomiędzy wybranymi palcami oraz jej kąta obrotu.
- Rozpoznawanie gestów dłoni w dwóch trybach: prostym (parę dostępnych gestów) oraz zaawansowanym, który rozoznaje alfabet języka migowego.
- Dostępność biblioteki poprzez platformę PyPi.

Stworzenie modułu będzie wymagało napisania klasy wykorzystującej odpowiednie biblioteki pozwalających na rozpoznanie elementów charakterystycznych dłoni oraz przetworzenia obrazu. Obraz będzie pochodził z kamery internetowej, który zostanie odpowiednio przetworzony z wykorzystaniem funkcji dostępnych

poprzez bibliotekę OpenCV. Przetworzony obraz zostanie wykorzystany przez metody biblioteki MediaPipe, która pozwoli na rozpoznanie elementów charakterystycznych dłoni. W napisanej klasie zostaną zaimplementowane metody, które pozwolą na rozpoznanie typu dłoni (prawa, lewa), odległości między paliczkiem palca wskazującego oraz kciuka, oraz obrotu dłoni.

Kolejnym elementem jest wygenerowanie modelu matematycznych klasyfikujących gesty dłoni. W tym celu zostanie wykorzystana biblioteka SciKit Learn wraz z dostępnymi poprzez nią algorytmami uczenia maszynowego. Odpowiednio zebrane dane pozwolą na przeprowadzenie procesu uczenia dla paru wybranych algorytmów.

Najważniejszym elementem pracy jest powyżej wymieniony ostatni punkt listy. Dzięki wykorzystaniu platformy PyPi deweloperzy będą mogli przy pomocy programu **pip** za pomocą jednej komendy zainstalować paczkę wraz ze wszystkimi wymaganymi zależnościami.

3.2 Podobne rozwiązania

Rozdział 4

Wymagania i narzędzia

- wymagania funkcjonalne i нефункционалне
- przypadki użycia (diagramy UML) - dla prac, w których mają zastosowanie
- opis narzędzi, metod eksperymentalnych, metod modelowania itp.
- metodyka pracy nad projektowaniem i implementacją - dla prac, w których ma to zastosowanie

4.1 Wymagania

Klasa powinna zawierać w sobie wszystkie niezbędne funkcje oraz parametry pozwalające na wykorzystanie jej w dowolnym projekcie. Modele rozpoznające gesty powinny zostać uprzednio przygotowane.

Pobranie modułu powinno być możliwe poprzez wykorzystanie standardowego programu do zarządzania paczkami **pip**. Ten menedżer pozwala na instalację paczki wraz z jej zależnościami, czyli innymi paczkami wymaganymi do porwanego działania pobieranego modułu.

4.2 Narzędzia

4.2.1 System operacyjny Ubuntu

Do stworzenia oprogramowania została wybrana dystrybucja Ubuntu systemu Linux. Dzięki takim elementom jak menedżer pakietów lub wbudowane wsparcie dla języka Python.

4.2.2 Python

Do napisanie biblioteki został wykorzystany język skryptowy Python. Ze względu na swoją budowę, Python nie należy do najbardziej wydajnych języków. Python jest językiem interpretowalnym, co oznacza, że jego program nie zostaje przekompilowany do kodu maszynowego, a zostaje on zinterpretowany przez program nazywany interpreterem. Pomimo niskiej wydajności języka Python, jest on dobrym narzędziem do tworzenia aplikacji wykorzystujących wizję komputerową oraz uczenia maszynowe. Głównie ze względu na dostępność oraz jakość bibliotek o wolnym źródle.

4.2.3 iPython

To interaktywna powłoka dla języka Python, która rozszerza jego działanie o introspekcję, czyli możliwość wykonywania poprzednich części programu zapisanych w komórkach. Dodatkowo IPython oferuje dodatkową składnię powłoki oraz wykonywanie komend wiersza poleceń.

4.2.4 Jupyter Notebook

Dodatkowym narzędziem jest Jupyter Notebook, który pozwala na wykonywanie części programu w osobnych komórkach. Pozwala to na łatwiejsze odnajdowanie błędów w programie. Jupyter Notebook został wykorzystany do napisania programu trenującego modele uczenia maszynowego rozpoznaące gesty. Dzięki strukturze takiego pliku, można w prosty sposób wykonywać program w danej komórce wiele razy. Na przykład pobieranie danych, dla różnych gestów.

4.2.5 Visual Studio Code - Środowisko Pracy

Całość klasy została napisana z wykorzystaniem edytora Visual Studio Code. Ten edytor został wybrany ze względu na jego wielozadaniowość. Visual Studio Code obsługuje jednocześnie programy napisane w Pythonie oraz pliki z rozszerzeniem **.ipynb**, czyli tak zwane zeszyty, które wykorzystują interaktywnego Pythona, czyli język iPython.

4.2.6 Platforma PyPi

Platforma PyPi jest platformą, na której udostępniane są moduły napisane w języku Python. Aktualnie jest ona zarządzana poprzez fundację "Python Software Foundation". Dzięki tej platformie, zupełnie za darmo można pobierać znajdujące się na niej oprogramowanie oraz udostępniać własne. Program **pip** w głównej mierze korzysta z PyPi jako głównego repozytorium z paczkami.

4.2.7 Programy **bdist_wheel** oraz **sdist**

Oba programy są niezbędne do przygotowania w pełni działającej paczki udostępnionej przez platformę PyPi. Program **sdist** pozwala na stworzenie źródłowej dystrybucji, czyli w praktyce pliku typu **.zip**, **.tar** czy też **.rar** oraz wielu innych, w których znajdują się wybrane pliki. Kolejnym programem jest **bdist_wheel**, który odpowiada za stworzenie uprzednio paczki typu **WHEEL**, która w przypadku wykorzystania plików napisanych w języku **C** uprzednio je kompiluje, dzięki czemu nie jest wymagany kompilator po stronie użytkownika. Ostatecznie, **WHEEL** pozwala na szybszą instalację paczki w porównaniu z budowaniem jej z plików źródłowych.

4.3 Metodyka Pracy

4.3.1 Praca z Jupyter Notebook

Rozdział 5

[Właściwy dla kierunku - np. Specyfikacja zewnętrzna]

5.1 Wymagania sprzętowe

5.1.1 Kamera

Elementem niezbędnym do korzystania z modułu OpenLeap jest kamera, która pozwoli na pozyskanie obrazu dłoni. Rozdzielczość matrycy kamery powinna pozwolić na rozpoznanie dłoni. Liczba klatek na sekundę powinna być wystarczająco wysoka, tak aby można było sprawnie korzystać z paczki. Nie ma tutaj minimalnych wymagań, lepsza kamera pozwoli jedynie na sprawniejsze działanie oprogramowania.

5.2 Wymagania programowe

Pierszym elementem jest język Python, który pozwoli na interpretację napisanego modułu. Reszta potrzebnych zależności, takich jak MediaPipe czy OpenCV zostanie zainstalowana automatycznie, jeśli moduł zostanie pobrany za pomocą programu pip.

5.3 Instalacja paczki

Instalacja paczki odbywa się poprzez wykorzystanie programu pip, który jest instalowany automatycznie razem z językiem Python. W zależności od wybranego systemu operacyjnego, komenda może przybierać inne formy.

```
1 $ pip install openleap
```

Informacje na temat biblioteki można znaleźć na platformie PyPi pod linkiem: <https://pypi.org/project/openleap/>. Na tej stronie znajduje się opis modułu, instrukcja instalacji oraz przykładowe programy i wykorzystania.

openleap 0.5.5 ✓ Latest version

`pip install openleap`

Released: less than 20 seconds ago

Hand tracking and gesture recognition module

Navigation

- [Project description](#)
- [Release history](#)
- [Download files](#)

Project links

- [Homepage](#)

Statistics

View statistics for this project via [Libraries.io](#), or by using [our public dataset on Google BigQuery](#)

Meta

License: MIT License (LICENSE)

Author: [Szymon Ciemala](#)

Maintainers

- [szymciem](#)

Classifiers

License

- [OSI Approved :: MIT License](#)

Operating System

- [OS Independent](#)

Project description

OpenLeap

Table of contents

- [OpenLeap](#)
 - [Table of contents](#)
 - [General Info](#)
 - [Technologies](#)
 - [Setup](#)
 - [Simple Example](#)
 - [Access Hand Information](#)
 - [Example](#)
 - [Another Example](#)

General Info

OpenLeap is an open source project that allows you to add hand gesture control to your Python projects.

Technologies

Project was created with technologies:

- Python
- OpenCV
- MediaPipe
- Scikit Learn

Setup

OpenLeap can be installed using pip, as shown below.

Rysunek 5.1: Strona modułu OpenLeap na PyPi

5.4 Przykładowy program

Paczkę można przetestować korzystając z wybudowanych gotowych funkcji. Program można napisać w powłoce języka Python lub po prostu zapisać go do

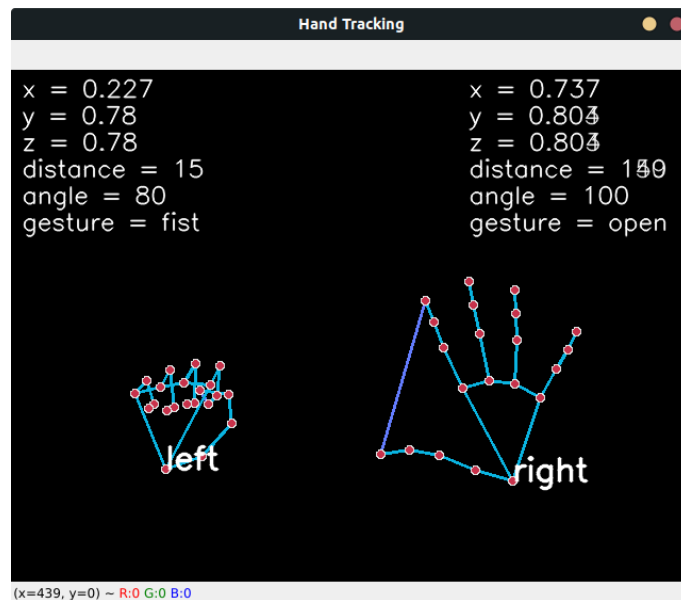
pliku.

```

1  import openleap
2
3  controller = openleap.OpenLeap(screen_show=True,
4                                  screen_type='BLACK',
5                                  show_data_on_image=True,
6                                  gesture_model='basic')
7
8  controller.loop()

```

Taki krótki program pozwoli na wyświetlenie okna z widocznymi dłońmi z oznaczonymi punktami charakterystycznymi wraz z opisem parametrów, takich jak obrót dłoni, jej pozycja czy rozpoznany gest. Zrzut ekranu testowego programu znajduje się poniżej.



Rysunek 5.2: Zrzut ekranu programu testowego

Dane zostają zapisane w liście składającej się z obiektów typu dataclass dla dłoni lewej i prawej, które przechowują wygenerowane informacje o danej dłoni. Operację pobrania danych można wykonać w poniżej przedstawiony sposób.

```

1  Zbierz dane

```

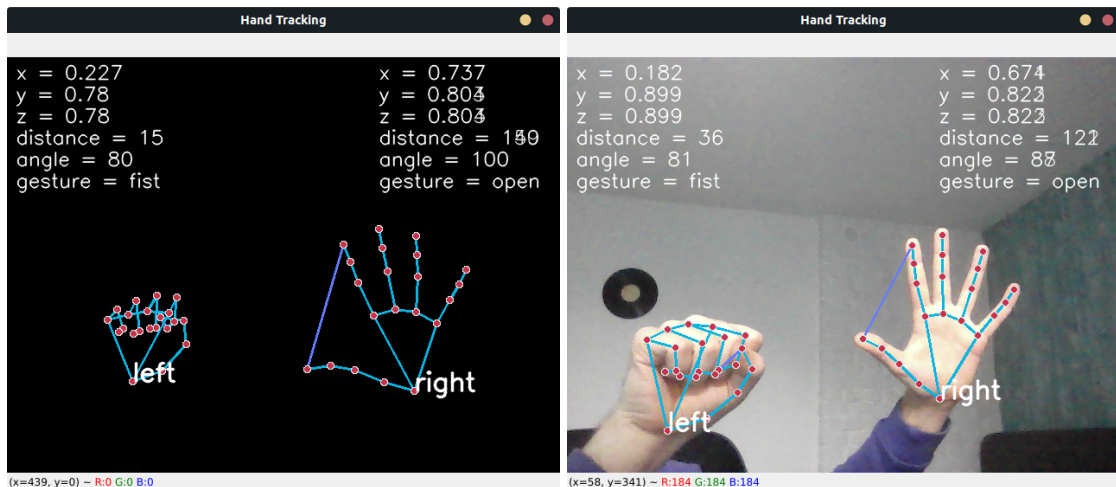
5.5 Funkcjonalność modułu

Paczkę można wykorzystać tam gdzie wymagane jest wykorzystanie gestów oraz ruchów dłoni. Paczka może znaleźć zastosowanie aplikacjach użytku codziennego, robotyce lub pełnić formę peryferium komputerowego.

5.5.1 Parametry

Obiekt kontrolera może zostać stworzony z wybranymi parametrami. Można wybrać czy należy wyświetlić podgląd kamery, czy wyświetlić dane na podglądzie oraz w konsoli oraz wybrać tło podglądu, czarne lub obraz z kamery.

1. **screen_show** - podgląd okna, typ **boolean**
2. **screen_type** - typ tła
 - (a) „CAM” - obraz z kamery
 - (b) „BLACK” - czarne tło
3. **show_data_in_console** - wyświetlanie danych w konsoli, typ **boolean**
4. **show_data_on_image** - wyświetlanie danych w oknie graficznym, typ **boolean**
5. **normalized_position** - wyświetlanie znormalizowanej pozycji, typ **boolean**
6. **gesture_model** - wybór modelu rozpoznającego gesty
 - (a) „basic” - gesty podstawowe
 - (b) „sign_language” - język migowy



(a) Obraz z kamery.

(b) Czarne tło.

Rysunek 5.3: Parametry wyświetlanego obrazu.

5.5.2 Dostępne funkcje

5.5.3 Główny wątek

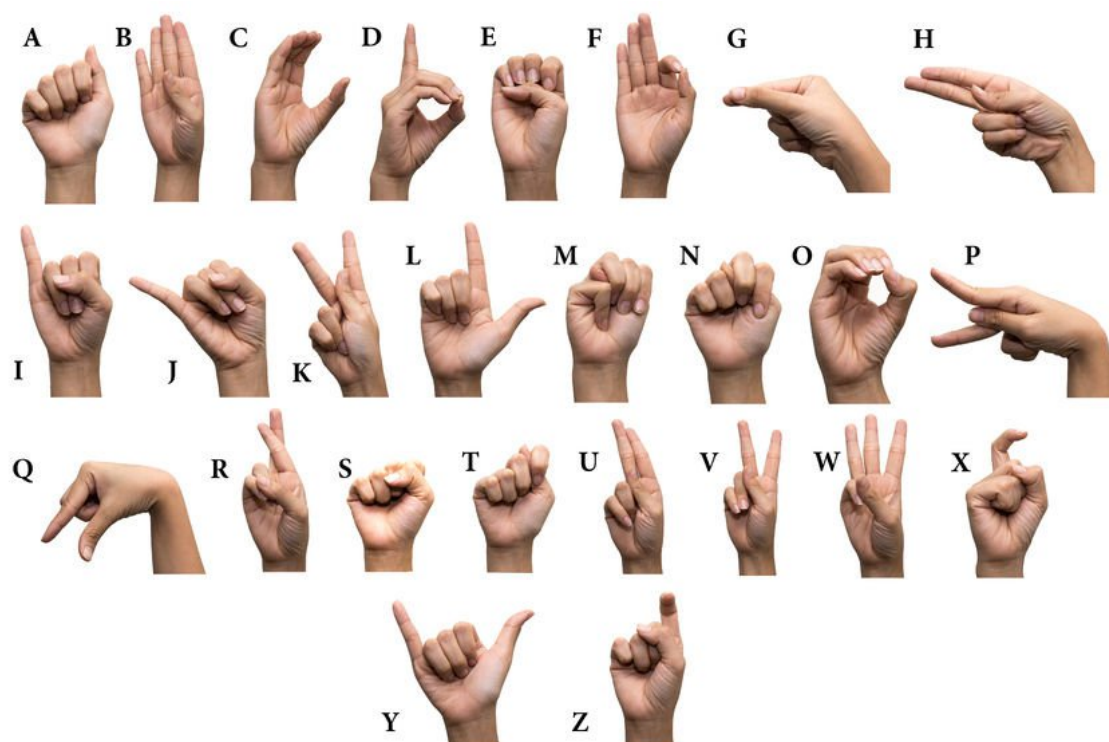
Działanie kontrolera opiera się na wywoływaniu funkcji **main()**. Tą funkcję można wywołać za pomocą wybudowanej funkcji **loop()**, która po prostu wywołuje funkcję **main()** w pętli wraz z warunkiem wyjścia z programu. Takie podejście może ułatwić stosowanie kontrolera w osobnym wątku aplikacji. Alternatywnym podejściem jest po prostu zastosowanie funkcji **main()** w pętli budowanego programu. Takie rozwiązanie daje podobny rezultat, ale bez konieczności wykorzystania osobnego wątku.

5.6 Przykłady użycia

5.6.1 Rozpoznawanie alfabetu w języku migowym

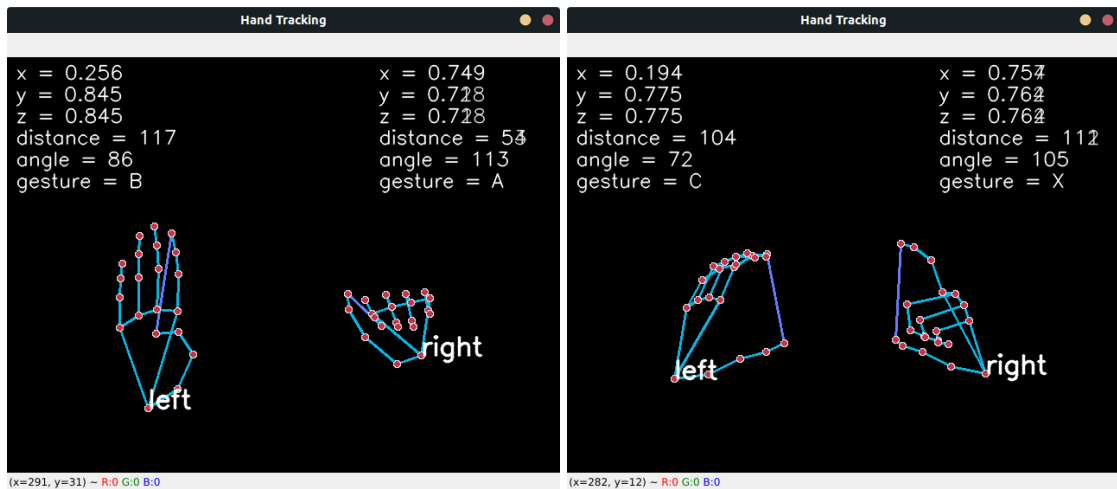
Pierwszym przykładem zastosowania jest wykorzystanie paczki do rozpoznawania alfabetu języka migowego. Taki program może umożliwić komunikację między osobą głuchoniemą posługującą się językiem migowym, a osobą, które ta-

kiego języka nie zna.



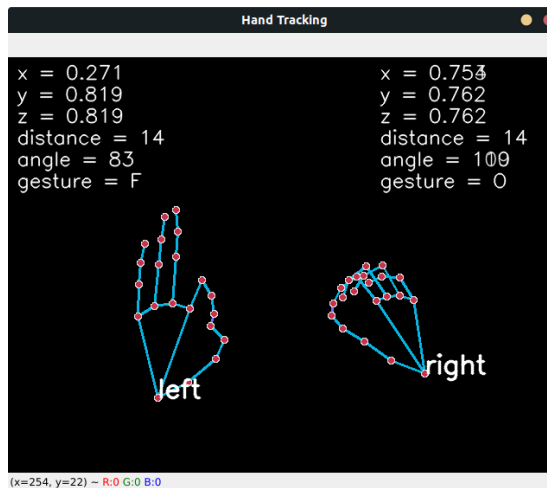
Rysunek 5.4: Gesty alfabetu języka migowego

Oprócz modelu rozpoznającego gesty alfabetu języka migowego, klasa została wyposażona w model rozpoznający podstawowe gesty, czyli gesty otwartej i zamkniętej dłoni. Wybór odpowiedniego modelu odbywa się w momencie tworzenia obiektu poprzez ustawienie parametru **gesture_model**.



(a) Litery A i B.

(b) Litery C i X.

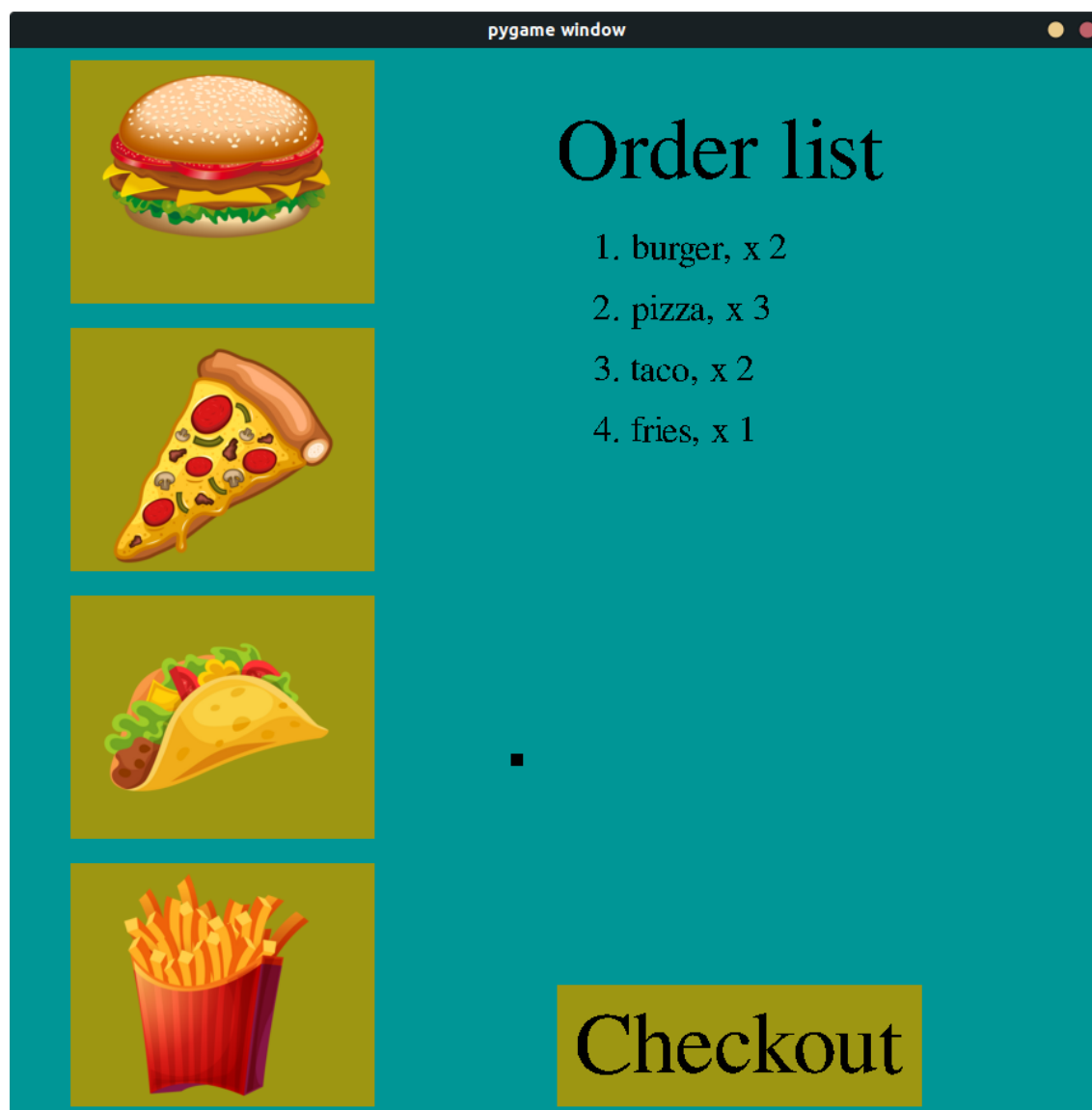


(c) Litery F i O.

Rysunek 5.5: Rozpoznawanie języka migowego .

5.6.2 Interaktywny Kiosk

Kolejnym przykładem jest interaktywny kiosk, który pozwala na złożenie zamówienia w sposób, który nie wymaga dotykania ekranu dotykowego. W dobie pandemii takie rozwiązanie może potencjalnie przyczynić się do spowolnienia rozprzestrzeniania się różnego rodzaju wirusów i drobnoustrojów.



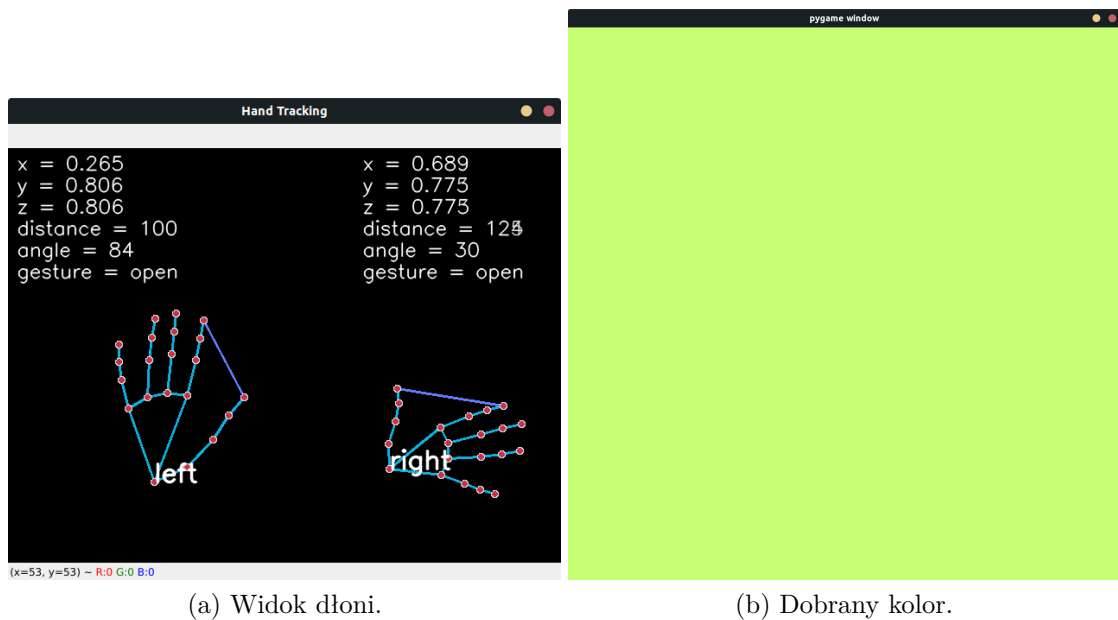
Rysunek 5.6: Zrzut ekranu kiosku interaktywnego

Wskaźnik kiosku interaktywnego jest sterowany poprzez pozycję dłoni. Kliknięcie w przycisk zostaje aktywowane za pomocą wykrycia odpowiednio małej odległości między końcówką palca wskazującego, a końcówką kciuka.

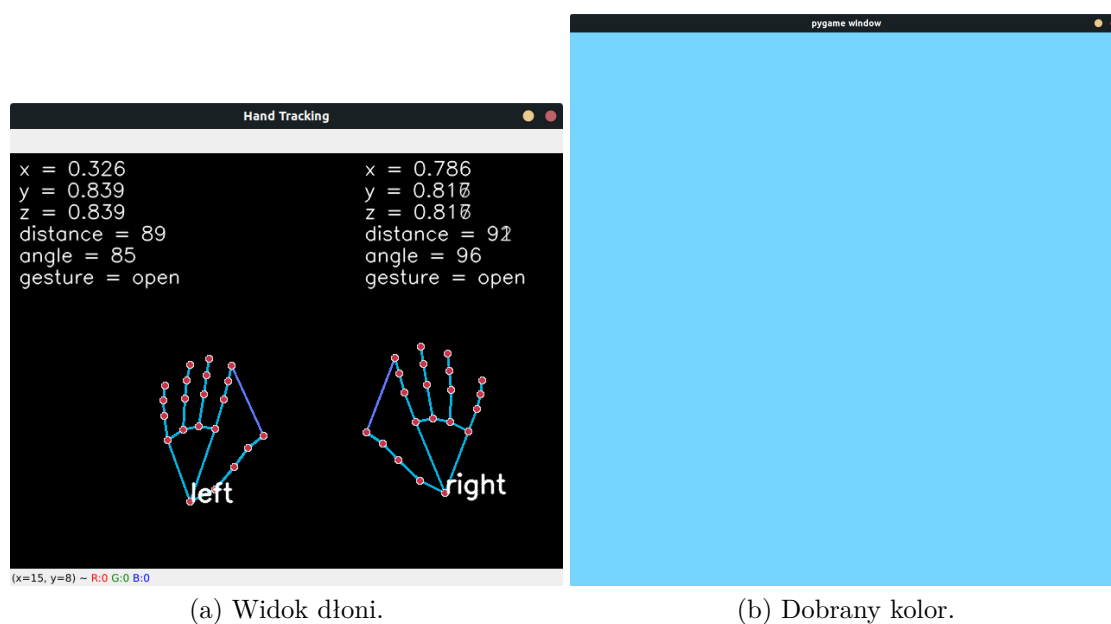
Odległość między tymi dwoma punktami jest obliczana automatycznie poprzez wbudowaną funkcję klasy OpenLeap.

5.6.3 Dobór koloru

Ostatnim przykładem jest wykorzystanie dłoni jako kontrolera, za którego pomocą można wybrać dowolny kolor. Takie zastosowanie może zostać wykorzystane w pracy grafika komputerowego. Dzięki temu użytkownik będzie mógł zmieniać kolor wykorzystywanego narzędzia bez przerywania pracy, na przykład malowania. Kolor można ustawiać tylko wtedy kiedy gest lewej ręki jest gestem otwartej dłoni.



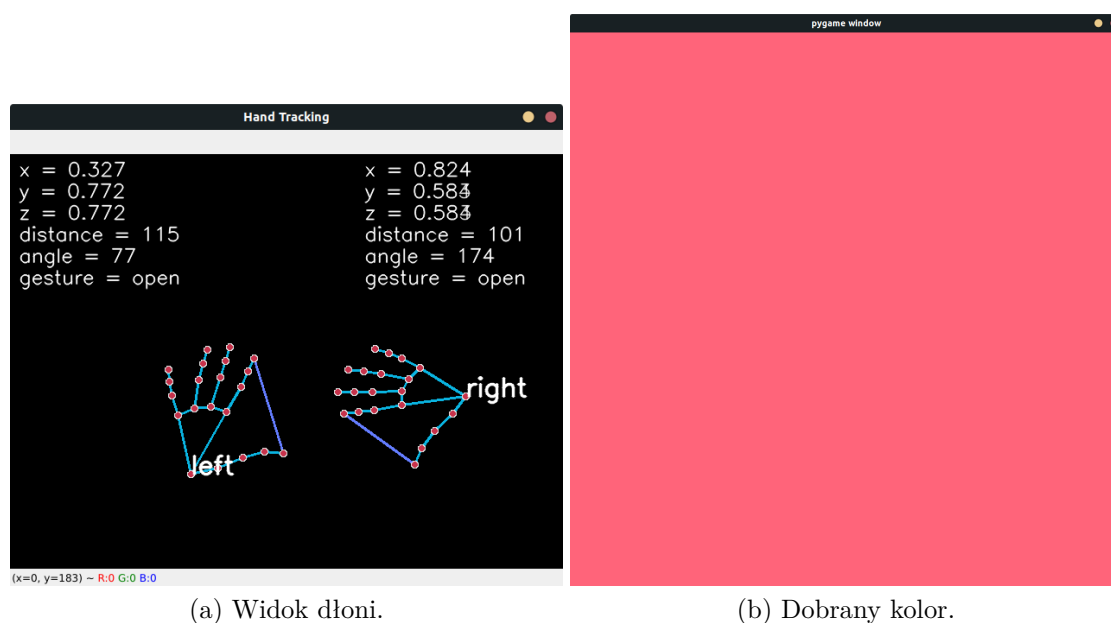
Rysunek 5.7: Obrót dłoni o 30 stopni.



(a) Widok dłoni.

(b) Dobrany kolor.

Rysunek 5.8: Obrót dłoni o 96 stopni.



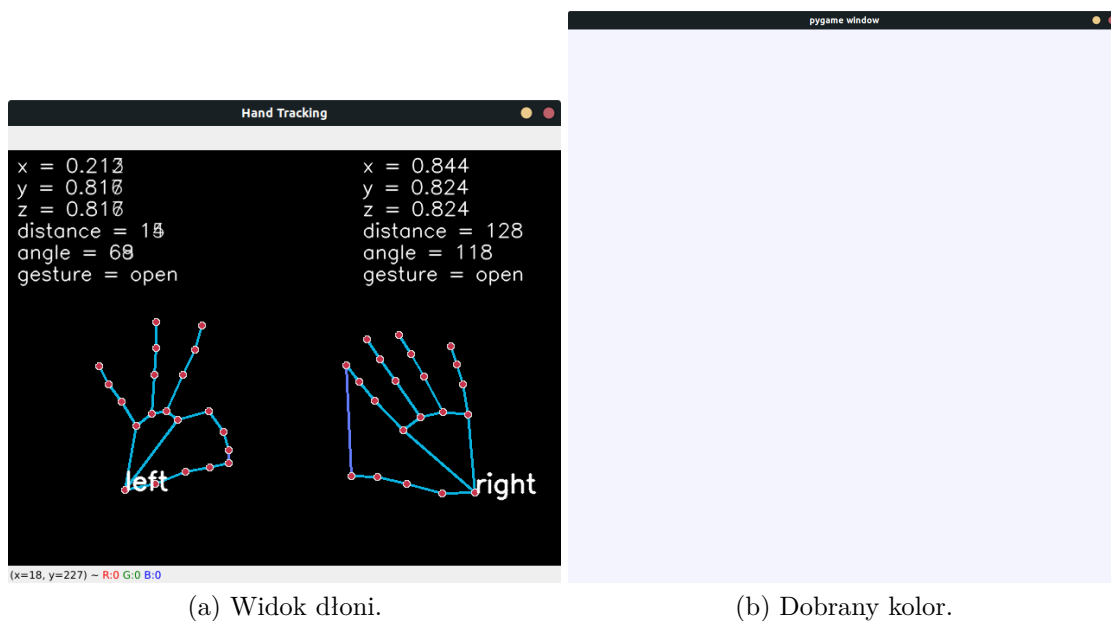
(a) Widok dłoni.

(b) Dobrany kolor.

Rysunek 5.9: Obrót dłoni o 174 stopni.

Dodatkową opcją jest ustawienie saturacji poprzez obliczenie jej wartości na

podstawie odległości między palcem wskazującym, a kciukiem.



Rysunek 5.10: Ustawienie saturacji

Podobnie jak obliczenie odległości między palcami, biblioteka oblicza obrót dłoni wokół osi Z punktu opisującego pozycję nadgarstka.

5.7 Stworzenie nowego modelu przez użytkownika

Rozdział 6

[Właściwy dla kierunku - np.Specyfikacja wewnętrzna]

Jeśli to Specyfikacja wewnętrzna:

- przedstawienie idei
- architektura systemu
- opis struktur danych (i organizacji baz danych)
- komponenty, moduły, biblioteki, przegląd ważniejszych klas (jeśli występują)
- przegląd ważniejszych algorytmów (jeśli występują)
- szczegóły implementacji wybranych fragmentów, zastosowane wzorce projektowe
- diagramy UML
- wymagania funkcjonalne i нефункционалне
- przypadki użycia (diagramy UML) - dla prac, w których mają zastosowanie
- opis narzędzi, metod eksperymentalnych, metod modelowania itp.
- metodyka pracy nad projektowaniem i implementacją - dla prac, w których ma to zastosowanie

6.1 Budowa klasy

Schemat UML itd

6.2 Struktury danych

6.2.1 Dataclass

Struktura typu **Dataclass** pozwala na stworzenie struktury podobnej do **struct** istniejącej w języku programowania **C**. Taka klasa pozwala na stworzenie obiektu składającego się jedynie z atrybutów wymaganych do opisu elementów klasy kontrolera. Dodatkowym atutem takiej klasy jest możliwość prostego odczytu zapisanych danych, korzystając z operatora kropki, a nie z operatów opisujących słownik lub listę.

```
1      @dataclass
2      class Data:
3          x : float = 0
4          y : float = 0
5          z : float = 0
6          distance: float = 0.0
7          angle: float = 0.0
8          gesture: str = None
9
```

W języku Python stworzenie klasy typu **dataclass** zaczyna się od zapisania dekoratora. W tym wypadku nie jest wymagana funkcja inicjalizująca obiekt, czyli `__init__`. Parametry początkowe można podać w chwili tworzenia obiektu, lub później. W programie zostały przypisane wartości początkowej, tak jak w przykładzie powyżej.

6.2.2 Słownik

Instancja stworzonego typu **Data** zostanie zainicjowana dla każdej dłoni (lewej oraz prawej). Te instancje zostaną zapisane w słowniku.

```
1      data = {'right':Data(), 'left':Data()}
2
```

Taka konstrukcja pozwoli użytkownikowi w prosty i przejrzysty sposób na odnajdowania potrzebnych wartości oraz informacji.

6.2.3 Pickle

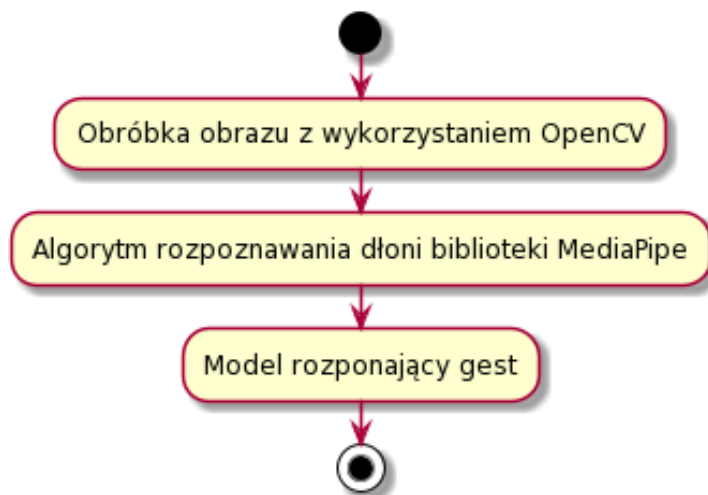
Modele matematyczne, które zostaną wytrenowane muszą zostać zapisane, tak aby można było je wykorzystać ponownie. Z tego powodu został wykorzystany plik typu **pickle**, który umożliwia zapis zmiennych oraz obiektów w postaci binarnej. Dzięki czemu można je wykorzystać ponownie pomimo restartu programu.

6.3 Rozpoznawanie dłoni

Pierwszym elementem projektu jest rozpoznanie dłoni poprzez wyznaczenie pozycji elementów charakterystycznych. Pozycja każdego z tych elementów, jak już zostało to opisane, jest względna według lewego górnego rogu obrazu kamery.

6.3.1 OpenCV - przygotowanie obrazu z kamery

Poprawne działanie modelu MediaPipe wymaga odpowiedniego przygotowania obrazu kamery. Działanie kontrolera odbywa się poprzez główną metodę **main()**.



Rysunek 6.1: Elementy charakterystyczne dłoni

Metoda **main()** jest główną funkcją, w której dokonywane są obliczenia oraz

przeszktałenia pozwalające na obliczenie obrotu dłonie, odległości między wybranymi palcami oraz na wykrycie gestu.

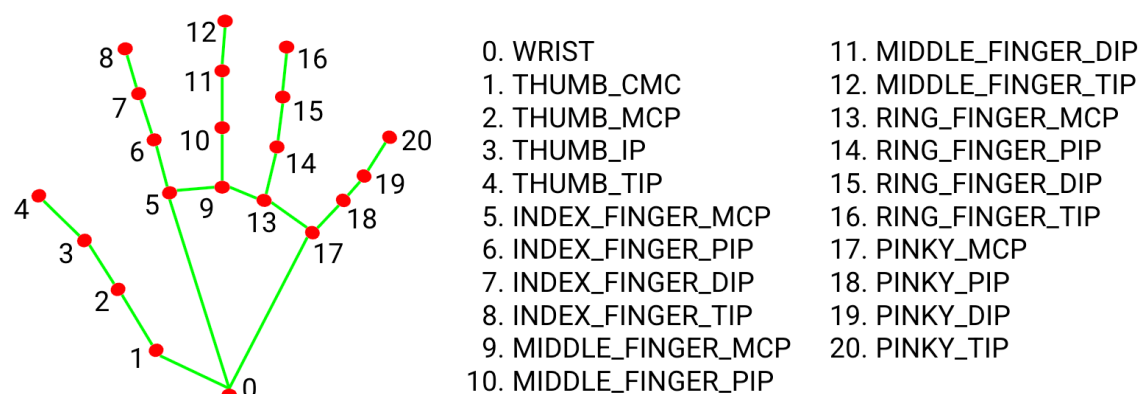
W pierwszym kroku tworzymy instancję klasy **VideoCapture** biblioteki **OpenCV**, która pozwoli na odczytywanie obrazu kamery.

6.3.2 MediaPipe - Elementy charakterystyczne

Biblioteka MediaPipe o otwartym źródle, udostępnia wieloplatformowe oraz konfigurowalne rozwiązania wykorzystujące uczenie maszynowe w dziedzinie rozpoznawania, segmentacji oraz klasyfikacji obiektów wizji komputerowej. Niektórymi z rozwiązań są:

- Rozpoznawanie twarzy
- Segmentacja włosów oraz twarzy
- Rozpoznawanie oraz określanie rozmiarów obiektów trójwymiarowych na podstawie obrazu dwuwymiarowego.

Model modułu MediaPipe pozwala na wyznaczenie pozycji 21 elementów charakterystycznych dłoni. Współrzędne X i Y są znormalizowane względem rozdzielczości obrazu kamery. Współrzędna X względem liczby pikseli w osi X, a współrzędna Y względem liczby pikseli w osi Y. Oś Z jest prostopadła do osi X i Y, z punktem początkowym w punkcie określającym pozycję nadgarstka. Współrzędna Z jest znormalizowana względem szerokości obrazu kamery, tak jak współrzędna X.



Rysunek 6.2: Elementy charakterystyczne dłoni

Pozycje nadgarstka, paliczków oraz stawów dłoni zostaną wykorzystane do obliczenia obrotu dłoni względem punktu 0 oraz do wytrenowania modeli uczenia maszynowego, których zadaniem będzie rozpoznawanie wybranych gestów.

6.3.3 Generowanie grafiki dłoni

Generowanie grafiki nałożonej na daną dłoń wykonuje się przy pomocy przygotowanej funkcji biblioteki MediaPipe, która współpracuje z OpenCV.

6.4 Pomiary oraz inne ważne elementy

6.4.1 Pozycja dłoni

6.4.2 Obrót dłoni

6.4.3 Odległość między palcami

6.5 Rozpoznawanie gestów

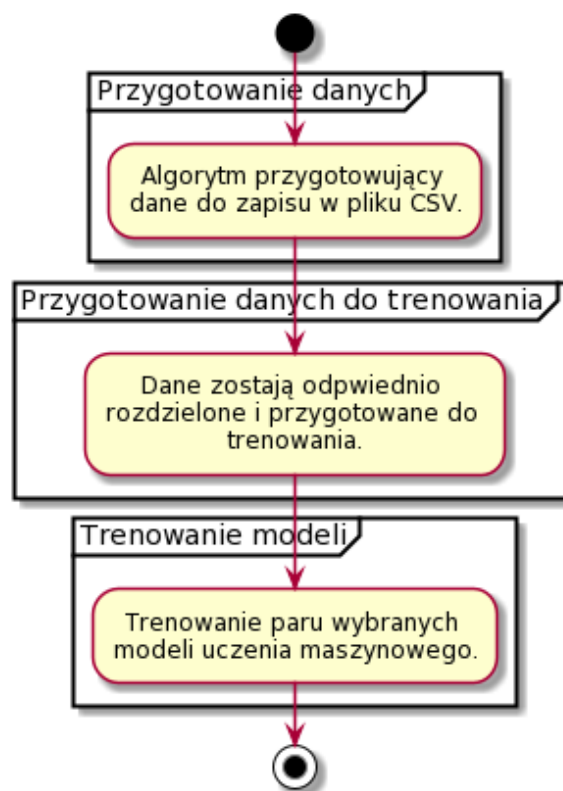
SciKit Learn to biblioteka, która oferuje różnego typu metody uczenia maszynowego. Biblioteka zawiera algorytmy klasyfikacji, regresji oraz analizy skupień. Przykładowym algorytmami w bibliotece są:

- Las losowy - polegająca na konstruowaniu wielu drzew decyzyjnych w czasie uczenia.
- Algorytm centroidów - algorytm wykorzystywany w analizie skupień.
- Maszyna wektorów nośnych - algorytm klasyfikujący, często wykorzystywany w procesie rozpoznawania obrazów.

O wszystkich dostępnych algorytmach informacje można znaleźć w ogólnodostępnej dokumentacji biblioteki.

6.5.1 Przygotowanie modeli uczenia maszynowego

Celem programu będzie stworzenie modeli matematycznych przy pomocy metod uczenia maszynowego, których celem będzie rozpoznawanie gestów dłoni. Proces tworzenie takiego modelu można podzielić na trzy kroki przedstawione na schemacie 6.3.



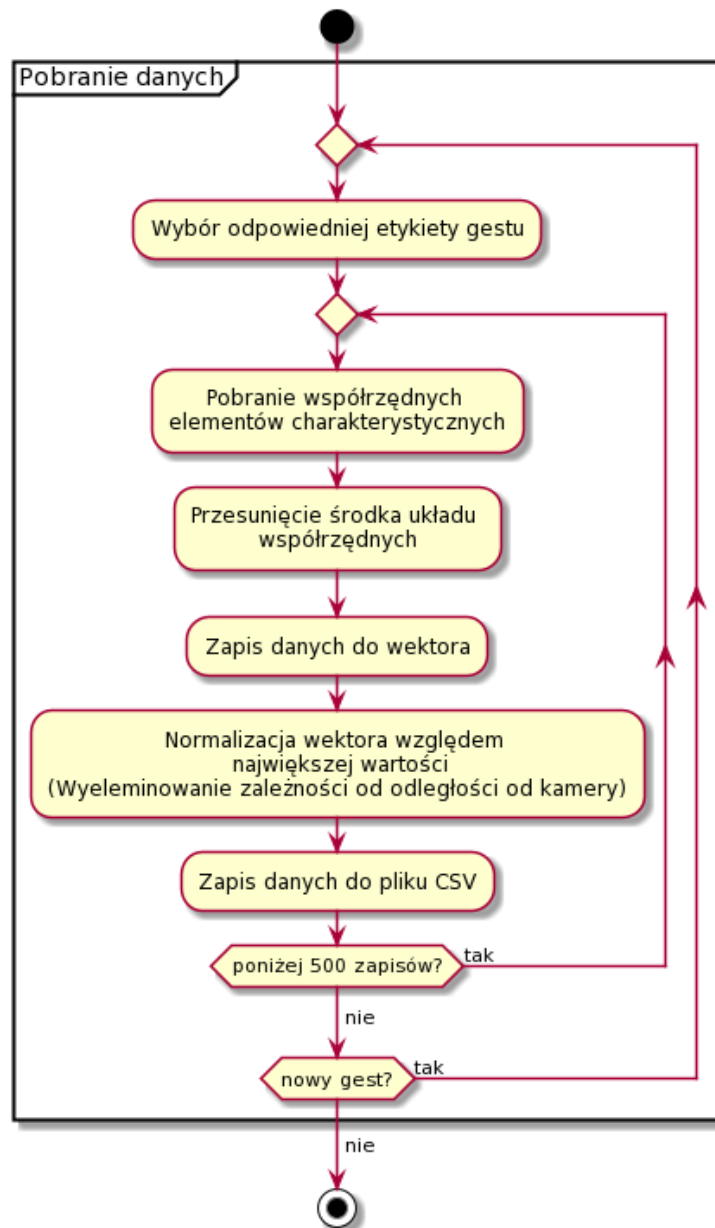
Rysunek 6.3: Ogólny algorytm przygotowania modeli uczenia maszynowego

Całość programu została napisana w notatniku Jupyter. Pozwala to na wykonanie pewnych części programu osobno, niezależnie od reszty programu. W praktyce każdy blok opisany w powyższym schemacie UML ma swoje odwzorowanie w notatniku. Na przykład, część zapisująca współrzędne do pliku będzie wykonywana tyle razy ile jest gestów do wytrenowania.

6.5.2 Zebranie danych

Algorytm zebrania danych polega na pobraniu współrzędnych oszacowanych przez model MediaPipe. Należy pamiętać o tym, że środek układu współrzędnych znajduje się lewym górnym rogu obrazu kamery. Oznacza to, że współrzędne w tej postaci nie nadają się do wyuczenia modelu matematycznego. Gest nie powinien być rozpoznawany na podstawie pozycji dłoni w obrazie lub jej odległości od

kamery. Należy się tych zależności pozbyć.



Rysunek 6.4: Ogólny algorytm przygotowania modeli uczenia maszynowego

Gest dłoni, można scharakteryzować na podstawie pozycji elementów dłoni. Podstawowym problemem jest fakt, że pozycje elementów charakterystycznych są

opisane względem układu współrzędnych, którego środek znajduje się w lewym górnym rogu obrazu pobranego z kamery. Pozycja dłoni na obrazie nie ma żadnego znaczenia w kwestii określania gestu. Poprawnie przygotowane pozycje elementów powinny zostać pozbawione zależności od pozycji dłoni względem obrazu. Przygotowanie danych do przetworzenia będzie wymagało paru operacji matematycznych. W takim wypadku należy przeprowadzić transformację, tutaj akurat przesunięcie układu współrzędnych do pozycji nadgarstka. Taka operacja pozwoli na opis pozycji elementów względem nadgarstka. Ostatecznie pozbywamy się pozycji nadgarstka z wektora danych, ponieważ jest ona środkiem nowego układu współrzędnych.

Macierz przesunięcia

$$M_p = \begin{bmatrix} 1 & 0 & 0 & -x_0 \\ 0 & 1 & 0 & -y_0 \\ 0 & 0 & 1 & -z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Tak naprawdę, przesunięcie układu współrzędnych w osi Z nie będzie miało miejsca. Wartości współrzędnej Z reszty elementów, oprócz elementu z indeksem równym 0, czyli nadgarstka, są określane właśnie nadgarstka. Dlatego też, macierz przesunięcia w osi Z jest równa 0.

$$M_p = \begin{bmatrix} 1 & 0 & 0 & -x_0 \\ 0 & 1 & 0 & -y_0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Aby dane mogły zostać zinterpretowane przez algorytmy uczenia maszynowego muszą one zostać przedstawione w postaci jednowymiarowej. Aktualna postać macierzy przedstawiającej współrzędne elementów charakterystycznych ma następującą postać. Indeksy współrzędnych są równoznaczne z indeksami elementów dłoni.

$$M_p = \begin{bmatrix} x'_1 & y'_1 & z'_1 \\ x'_2 & y'_2 & z'_2 \\ \vdots & \vdots & \vdots \\ x'_{21} & y'_{21} & z'_{21} \end{bmatrix}$$

Dane w postaci jednowymiarowej mają postać następującego wektora.

$$A_f = [x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ \cdots \ y_{21} \ z_{21}]$$

Drugim krokiem jest uniezależnienie pozycji elementów od odległości dłoni od kamery. Najprostszym rozwiązaniem jest normalizacja wektora danych względem największej bezwzględnej wartości.

Normalizacja

$$A_n = \frac{A_f}{\max(\text{abs}(A_f))}$$

Każdy nowy wektor zostaje zapisany do pliku CSV z odpowiednią etykietą. Zebrane dane posłużą do wytrenowania algorytmów uczenia maszynowego.

6.5.3 Budowa pliku CSV

Dane zapisane w pliku CSV opisują przykładowe współrzędne wszystkich elementów dłoni wraz z przypisaną etykietą oznaczającą gest.

6.5.4 Metody klasyfikacji - uczenie maszynowe

Przygotowane dane zostają odczytane z pliku CSV. W pierwszym kroku należy rozdzielić je na dwie części: współrzędne (dane wejściowe) oraz etykiety (dane wyjściowe). W kolejnym kroku należy te dwie grupy podzielić na grupę trenującą i grupę testową. Zadaniem grupy testowej będzie trenowanie wybranych modeli matematycznych, a grupy testowej przetestowanie ich dokładności.



Rysunek 6.5: Ogólny algorytm przygotowania modeli uczenia maszynowego

W celu wybrania najlepszej metody klasyfikacji, zostnie wybranych kilka algorytmów. Każdy z nich stworzy swój model, a ostatecznie zostanie sprawdzona ich poprawności z wykorzystaniem grupy testowej. Model z najlepszym wynikiem zostanie zapisany do pliku typu **pickle**. W języku Python pliki typu **pickle** pozwalają na zapis zmiennych, obiektów lub innych struktur danych, które mają zostać wykorzystane w po zakończeniu programu.

6.5.5 Wybrane algorytmy klasyfikujące

Do wytrenowania modeli uczenia maszynowego z biblioteki SciKit Learn zostały wybrane następujące algorytmy.

- Logistic Regression

- Nearest Centroid
- Decision Tree Classifier
- Random Forest Classifier
- SGD Classifier
- Gradient Boosting Classifier
- MPL Classifier

6.5.6 Badanie dokładności każdego z algorytmów

6.5.7 Ponowne wykorzystanie modelu

Gotowy model pobieramy i testujemy w przykładowym programie.

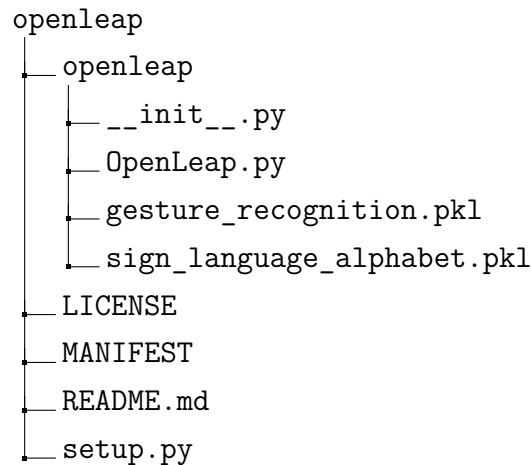
6.6 Paczka PyPi

6.6.1 Budowa paczki

Ostatecznym krokiem jest przygotowanie programu w formie paczki, która zostanie udostępniona na platformie PyPi. Wymaga to przygotowania odpowiednich plików konfiguracyjnych oraz zastosowania stosownych narzędzi do stworzenia pliku **wheel** oraz **tar**.

6.6.2 Struktura Paczki

Pierwszym krokiem jest przygotowanie odpowiedniej struktury paczki. Do tego celu został stworzony folder o poniższej strukturze. W tym folderze znajdują się wszystkie potrzebne elementy paczki. W podfolderze o tej samej nazwie znajduje się główna część modułu, czyli plik **.py**, w którym zapisana jest klasa **OpenLeap**. Dodatkowo w tym folderze znajdują się pliki typu **pickle**, w których zapisane są modele rozpoznające gesty.



Rysunek 6.6: Struktura paczki PyPi

6.6.3 Pliki Konfiguracyjne

Pliki `setup.py` oraz `MANIFEST` są plikami, które odpowiadają za konfigurację oraz opis paczki. W pliku `setup.py` zapisany jest numer aktualnej wersji, autor, kontakt do autora, nazwa paczki itp.

6.6.4 Załadowanie paczki do repozytorium

Przed załadowaniem paczki do repozytorium, należy stworzyć zapakowaną paczkę źródłową, na przykład typu `.tar` oraz plik typu `WHEEL`. Oba pliki spełniają tę samą funkcję, czyli przechowywanie niezbędnych elementów paczki oraz umożliwiają ich instalację na systemie użytkownika. Plik `WHEEL` pozwala na dużo szybszy proces instalacji niż instalacja ze źródła, czyli paczki typu `.tar`.

Rozdział 7

Weryfikacja i walidacja

- sposób testowania w ramach pracy (np. odniesienie do modelu V)
- organizacja eksperymentów
- przypadki testowe zakres testowania (pełny/niepełny)
- wykryte i usunięte błędy
- opcjonalnie wyniki badań eksperymentalnych

Tablica 7.1: Opis tabeli nad nią.

ζ	metoda						
	alg. 1	alg. 2	alg. 3			alg. 4, $\gamma = 2$	
			$\alpha = 1.5$	$\alpha = 2$	$\alpha = 3$	$\beta = 0.1$	$\beta = -0.1$
0	8.3250	1.45305	7.5791	14.8517	20.0028	1.16396	1.1365
5	0.6111	2.27126	6.9952	13.8560	18.6064	1.18659	1.1630
10	11.6126	2.69218	6.2520	12.5202	16.8278	1.23180	1.2045
15	0.5665	2.95046	5.7753	11.4588	15.4837	1.25131	1.2614
20	15.8728	3.07225	5.3071	10.3935	13.8738	1.25307	1.2217
25	0.9791	3.19034	5.4575	9.9533	13.0721	1.27104	1.2640
30	2.0228	3.27474	5.7461	9.7164	12.2637	1.33404	1.3209
35	13.4210	3.36086	6.6735	10.0442	12.0270	1.35385	1.3059
40	13.2226	3.36420	7.7248	10.4495	12.0379	1.34919	1.2768
45	12.8445	3.47436	8.5539	10.8552	12.2773	1.42303	1.4362
50	12.9245	3.58228	9.2702	11.2183	12.3990	1.40922	1.3724

Rozdział 8

Podsumowanie i wnioski

- uzyskane wyniki w świetle postawionych celów i zdefiniowanych wyżej wymagań
- kierunki ewentualnych danych prac (rozbudowa funkcjonalna ...)
- problemy napotkane w trakcie pracy

Dodatki

Spis skrótów i symboli

DNA kwas deoksyrybonukleinowy (ang. *deoxyribonucleic acid*)

MVC model – widok – kontroler (ang. *model-view-controller*)

N liczebność zbioru danych

μ stopień przyleżności do zbioru

\mathbb{E} zbiór krawędzi grafu

\mathcal{L} transformata Laplace’a

Źródła

Jeżeli w pracy konieczne jest umieszczenie długich fragmentów kodu źródłowego, należy je przenieść do załącznika.

Zawartość dołączonej płyty

Do pracy dołączona jest płyta CD z następującą zawartością:

- praca (źródła $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ owe i końcowa wersja w pdf),
- źródła programu,
- dane testowe.

Spis rysunków

5.1	Strona modułu OpenLeap na PyPi	15
5.2	Zrzut ekranu programu testowego	16
5.3	Parametry wyświetlanego obrazu.	18
5.4	Gesty alfabetu języka migowego	19
5.5	Rozpoznawanie języka migowego	20
5.6	Zrzut ekranu kiosku interaktywnego	21
5.7	Obrót dłoni o 30 stopni.	22
5.8	Obrót dłoni o 96 stopni.	23
5.9	Obrót dłoni o 174 stopni.	23
5.10	Ustawienie saturacji	24
6.1	Elementy charakterystyczne dłoni	27
6.2	Elementy charakterystyczne dłoni	29
6.3	Ogólny algorytm przygotowania modeli uczenia maszynowego	31
6.4	Ogólny algorytm przygotowania modeli uczenia maszynowego	32
6.5	Ogólny algorytm przygotowania modeli uczenia maszynowego	35
6.6	Struktura paczki PyPi	37

Spis tablic

7.1	Opis tabeli nad nią.	40
-----	------------------------------	----