

Rozpoznawanie człowieka metodami biometrii

Projekt 3. — Rozpoznawanie na podstawie twarzy

Raport

Bartłomiej Dach

12 maja 2019

Poniższy dokument stanowi sprawozdanie z implementacji aplikacji dokonującej rozpoznawania człowieka na podstawie zdjęć twarzy w pozycji frontальной. W dokumencie opisano porównywane metody oraz zawarto wyniki eksperymentalne dla obu metod na dostarczonym zbiorze zdjęć.

1 Wstęp

Twarz to cecha biometryczna zaliczająca się do cech biologicznych. Większość mierzalnych cech twarzy (geometria, kolor skóry) jest opartych na cechach anatomicznych lub etnicznych. Z drugiej strony aspekty takie, jak zarost czy wiek utrudniają wykorzystanie tych cech, ponieważ zmieniają się z czasem.

Rozpoznawanie na podstawie twarzy jest częścią coraz większej liczby urządzeń konsumenckich, szczególnie w branży urządzeń mobilnych. Ponieważ akwizycja obrazu twarzy nie wymaga bezpośredniego kontaktu, a coraz więcej smartfonów posiada kamery po przedniej stronie, jest to wygodna metoda potwierdzenia tożsamości użytkownika telefonu.

W ramach projektu zaimplementowano dwie metody oparte na identyfikacji punktów charakterystycznych twarzy. W obu przypadkach do znajdowania twarzy na zdjęciach i lokalizacji punktów charakterystycznych wykorzystano bibliotekę `dlib`. Metody różnią się właściwym przetwarzaniem punktów — w jednej metodzie punkty charakterystyczne wykorzystywane są bezpośrednio po pewnych korekcyjnych przekształceniach geometrycznych, zaś w drugiej w celu porównania z rozwiązaniami *state-of-the-art* wykorzystano wytrenowany klasyfikator ResNet [4, 7]. Dokładniejsze informacje znajdują się w podrozdziale 3.3.

2 Opis aplikacji

Opracowana aplikacja została zaimplementowana w języku skryptowym Python w wersji 3.5.2. Wybór umotywowany był minimalizacją czasu implementacji poprzez wykorzystanie istniejących bibliotek *open source*. Aplikacja ma postać zbioru skryptów konsolowych — zdecydowano się na rezygnację z interfejsu graficznego ze względu na jego niewielką przydatność w stosunku do czasu implementacji. Sposób wywołania skryptów opisano w podrozdziale 2.2.

2.1 Zastosowane biblioteki

W implementacji zastosowano zbiór bibliotek *open source* ułatwiających implementację systemu. Pełna lista zastosowanych bibliotek znajduje się w tabeli 1.

Nr	Nazwa	Opis	Licencja	
1	<code>dlib</code> 19.17.0	Biblioteka wspomagająca proces rozpoznawania twarzy	Boost	[6]
2	<code>matplotlib</code> 3.0.3	Tworzenie wykresów i wizualizacji	PSF	[5]
3	<code>pandas</code> 0.24.2	Struktury do manipulacji i analizy danych	BSD	[8]
4	<code>seaborn</code> 0.9.0	Rozszerzone wizualizacje danych	BSD	[11]
5	<code>tqdm</code> 4.31.1	Biblioteka wspomagająca do pasków postępu w skryptach	MPL	[2]

Tablica 1: Lista bibliotek użytych w projekcie

2.2 Instrukcja obsługi

W celu uruchomienia skryptów do rozpoznawania konieczne jest zainstalowanie interpretera Python oraz bibliotek zawartych w tabeli 1. Aby zainstalować wymagane biblioteki, należy wywołać polecenie

```
$ pip3 install -r requirements.txt
```

gdzie plik `requirements.txt` to plik dołączony do źródeł aplikacji

2.2.1 Skrypt `recognize.py`

Skrypt `recognize.py` pozwala na sprawdzenie, czy twarze na podanym zdjęciu należą do osób z podanego zbioru treningowego. Wykonywanie skryptu powinno być zgodne ze składnią:

```
recognize.py [-h] (-d TRAINING_SET_DIR | -f EMBEDDING_FILE)
              [-l LANDMARK_MODEL] (-n | -r RECOGNITION_MODEL)
              [-t THRESHOLD] [-o OUTPUT_EMBEDDINGS]
              TEST_IMAGE
```

gdzie poszczególne argumenty to:

- `-h` — wyświetla pomoc,
- `-d TRAINING_SET_DIR` — oznacza, że wektory cech dla zdjęć ze zbioru testowego powinny być obliczone na podstawie zdjęć z podanego folderu `TRAINING_SET_DIR`. Folder ze zdjęciami powinien zawierać plik `labels.csv`, zawierającego etykietę przedstawionej osoby dla każdego zdjęcia ze zbioru.
- `-f EMBEDDING_FILE` — oznacza, że wektory cech dla zdjęć ze zbioru testowego powinny być wczytane z istniejącego pliku `EMBEDDING_FILE`,
- `-l LANDMARK_MODEL` — zawiera ścieżkę do pliku zawierającego model używany do lokalizacji punktów charakterystycznych twarzy. W przypadku braku tego parametru domyślnie zachodzi próba użycia pliku o nazwie
`shape_predictor_68_face_landmarks.dat`
z katalogu roboczego.
- `-n` — oznacza, że do rozpoznawania ma być zastosowany własny algorytm normalizacji punktów charakterystycznych,
- `-r RECOGNITION_MODEL` — oznacza, że do rozpoznawania ma być zastosowany wytrenowany model ResNet załadowany z pliku `RECOGNITION_MODEL`,
- `-t` — oznacza próg liczbowy, którego przekroczenie powoduje odrzucenie zdjęcia ze zbioru testowego przy klasyfikacji twarzy,
- `-o OUTPUT_EMBEDDINGS` — pozwala na zapisanie wektorów cech po wyznaczeniu punktów charakterystycznych do pliku,
- `TEST_IMAGE` — ścieżka do zdjęcia z twarzami do identyfikacji.

2.2.2 Skrypt `confusion_verification.py`

Skrypt `confusion_verification.py` dokonuje testu wybranego klasyfikatora dla zadania weryfikacji tożsamości. Składnia wywołania powinna być zgodna z następującym wzorcem:

```
confusion_verification.py [-h]
                           (-d TRAINING_SET_DIR | -f EMBEDDING_FILE)
                           [-l LANDMARK_MODEL]
                           (-n | -r RECOGNITION_MODEL)
                           [-o OUT_FILE_PREFIX]
```

Wszystkie parametry poza `-o` mają to samo znaczenie, co w przypadku skryptu `recognize.py`. Flaga `-o` oznacza w tym przypadku prefiks dla plików tworzonych przez skrypt.

Wynikiem skryptu są dwa pliki:

1. Pierwszy plik o sufiksie `distance_matrix.csv` zawiera symetryczną macierz odległości Euklidesa między wektorami cech wyznaczonymi przez klasyfikator na podstawie punktów charakterystycznych.
2. Drugi plik o sufiksie `error_rates.csv` zawiera stosunki błędów w zadaniu weryfikacji w zależności od przyjętego progu. Dla każdej pary obrazów klasyfikator uznaje dwa obrazy za obrazy tej samej osoby, jeśli odległość Euklidesa między ich wektorami cech jest mniejsza niż przyjęty próg. Na tej podstawie obliczane są wskaźniki FAR i FRR.

2.2.3 Skrypt `confusion_identification.py`

Skrypt `confusion_identification.py` dokonuje testu wybranego klasyfikatora dla zadania identyfikacji osoby na zdjęciu. Składnia wywołania powinna być zgodna z następującym wzorcem:

```
confusion_identification.py [-h]
                           (-d TRAINING_SET_DIR | -f EMBEDDING_FILE)
                           [-l LANDMARK_MODEL]
                           (-n | -r RECOGNITION_MODEL)
                           [-o OUT_FILE_PREFIX]
```

gdzie parametry są interpretowane tak samo, jak w skrypcie `confusion_identification.py`. Wynikiem skryptu jest pojedynczy plik o sufiksie `confusion_matrix.csv`, zawierający macierz pomyłek obliczoną dla klasyfikatora. Wiersze macierzy oznaczają etykietę osoby na zdjęciu (wzorcową), zaś kolumny — etykietę zwróconą przez klasyfikator.

3 Opis metody

Proces rozpoznawania na podstawie zdjęcia można podzielić na dwa podstawowe etapy. Pierwszy z nich zajmuje się lokalizacją twarzy na obrazie, poprzez określenie prostokątnego obszaru, w którym zawierają się znalezione twarze. Drugi zaś zajmuje się faktycznym rozpoznawaniem.

Istnieje wiele podejść do rozpoznawania; część z nich, jak np. metoda twarzy własnych, operuje bezpośrednio na znalezionym fragmencie obrazu z twarzą, podczas gdy inne opierają się na lokalizacji i przetwarzaniu tzw. punktów charakterystycznych, czyli ustalonych punktów związanych z anatomią i geometrią twarzy. W ramach projektu skorzystano z tego drugiego podejścia.

Po wyznaczeniu punktów charakterystycznych następuje proces ich przekształcenia na cechy, które wykorzystywane są przez końcowy klasyfikator. W projekcie porównano dwa podejścia, opisane szerzej w podrozdziale 3.3.

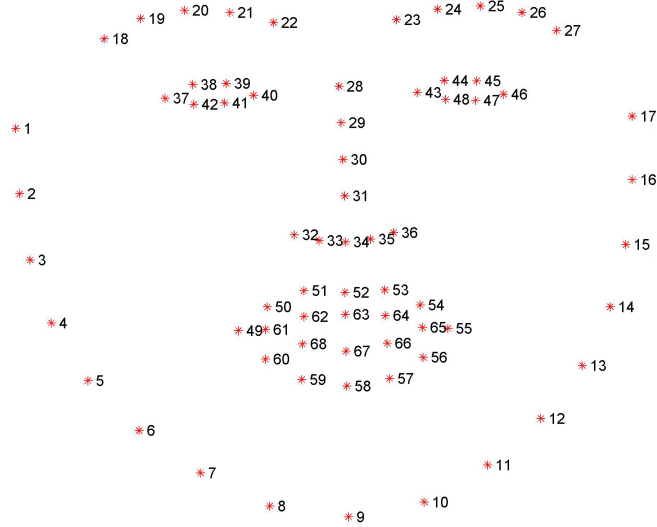
3.1 Rozpoznawanie twarzy na zdjęciu

Do implementacji etapu rozpoznawania twarzy wykorzystane zostały gotowe elementy biblioteki `dlib`. Wbudowany w bibliotekę detektor twarzy w postaci frontalnej wykorzystuje metodę histogramu zorientowanych gradientów (ang. *Histogram of Oriented Gradients*, HOG) [3].

Ta metoda po wstępnej normalizacji i korekcji gamma oblicza gradient obrazu poprzez zastosowanie filtrów splotowych do wykrywania krawędzi. Następnie gradienty w poszczególnych punktach są gromadzone w tzw. komórki, w ramach których następuje ważone uśrednianie. Później, aby zminimalizować wpływ oświetlenia, wykonywana jest lokalna normalizacja kontrastu. Na koniec dwuwymiarowe bloki są okienkowane (wycinane są fragmenty obrazu), i wynikowe okna są podawane do klasyfikatora, decydującego, czy układ gradientów w ramach okna jest zgodny z szukanym typem obiektu — w tym przypadku, frontalnym zdjęciem twarzy.

3.2 Lokalizacja punktów charakterystycznych

Po zlokalizowaniu prostokata ograniczającego twarz, następuje lokalizacja punktów charakterystycznych. Istnieje wiele propozycji zbiorów punktów wykorzystywanych przy identyfikacji; w tym przypadku zdecydowano się zastosować układ 68 punktów przedstawionych na rysunku 1.



Rysunek 1: Wizualizacja 68 punktów charakterystycznych wykorzystywanych do rozpoznawania twarzy, wytypowanych przez *Intelligent Behaviour Understanding Group*. Źródło: [10]

W projekcie zastosowano gotowy model predykcyjny wyznaczający punkty charakterystyczne, wytrenowany na ręcznie oetykietowanym zbiorze danych ibug 300-W [7]. Lokalizuje on punkty charakterystyczne dla znalezionych twarzy, które następnie wykorzystywane są do ekstrakcji cech.

3.3 Klasyfikacja twarzy na podstawie punktów charakterystycznych

Przy klasyfikacji zastosowano dwa podejścia. Pierwsze, nazwane odtąd normalizacyjnym, wykorzystuje wyliczone punkty charakterystyczne, próbując wyeliminować wpływ obrotów i skalowania na działanie klasyfikatora. Z kolei drugie opiera się na sieciach neuronowych opartych na architekturze ResNet [4].

3.3.1 Podejście normalizacyjne

Przy podejściu normalizacyjnym celem jest wyeliminowanie rotacji i skalowania. W celu wyeliminowania rotacji założono referencyjnie, że linia od lewego kącika lewego oka (punkt nr 37) do prawego kącika prawego oka (punkt nr 46) powinna być pozioma, zaś grzbiet nosa (linia od punktu 28 do punktu 34) — pionowa.

Aby to osiągnąć, zastosowano dwie operacje ścinania wzdłuż obu osi obrazu, przeprowadzające i -ty punkt charakterystyczny (x_i, y_i) ($i = 1, \dots, 68$) na punkt (x'_i, y'_i) , którego

współrzędne są określone wzorami

$$\begin{aligned}x'_i &= x_i - y_i \cdot \frac{x_{34} - x_{27}}{y_{34} - y_{27}} \\y'_i &= y_i - x_i \cdot \frac{y_{46} - y_{37}}{x_{46} - x_{37}}\end{aligned}$$

Następnie w celu wyeliminowania wpływu skalowania punkty charakterystyczne zostały przeprowadzone na kwadrat $[-0.5, 0.5] \times [-0.5, 0.5]$, zgodnie z wzorami

$$\begin{aligned}x''_i &= \frac{x'_i - x'_{\min}}{x'_{\max} - x'_{\min}} - 0.5 \\y''_i &= -\frac{y'_i - y'_{\min}}{y'_{\max} - y'_{\min}} + 0.5\end{aligned}$$

Zmiana znaku na współrzędnej Y uzasadniona jest tym, że piksele obrazów są indeksowane odwrotnie niż przyjmuje konwencja matematyczna (tj. współrzędna Y piksela rośnie w dół obrazu). Po tych dwóch operacji współrzędne X i Y obrazu są traktowane jako cechy do klasyfikacji — to podejście skutkuje więc uzyskaniem wektora złożonego z $2 \cdot 68 = 136$ cech. Warto zauważyć, że podejście to ignoruje barwy obrazu, lecz korzysta wyłącznie z geometrycznych cech twarzy.

3.3.2 Wykorzystanie klasyfikatora ResNet

Dla porównania jako drugą metodę wykorzystano wytrenowany wcześniej klasyfikator dostarczony przez twórców biblioteki `dlib`. Tym klasyfikatorem jest sieć neuronowa o zmodyfikowanym modelu ResNet [4], o 29 warstwach konwolucyjnych, wytrenowana na zbiorze danych składającym się z ok. 3 milionów twarzy. Klasyfikator transformuje punkty charakterystyczne na wektor w 128-wymiarowej przestrzeni. Funkcja błędu była skonstruowana tak, aby punkty korespondujące z zdjęciami twarzy jednej osoby stanowiły rozłączne kule o promieniu 0.6 [7].

3.3.3 Finalna klasyfikacja

W przypadku obu klasyfikatorów przyjęto najprostszą klasyfikację wyjściowych wektorów — klasyfikator 1-NN. Klasyfikacja polega więc na wyborze wektora reprezentującego twarz ze zbioru referencyjnego o najmniejszej odległości w metryce Euklidesowej od wektora wyznaczonego dla identyfikowanej twarzy.

4 Wyniki eksperymentalne

W ramach eksperymentu oba zaimplementowane klasyfikatory porównano pod względem efektywności w zadaniach weryfikacji i identyfikacji tożsamości osób. Testy wykonano

na zbiorze 450 zdjęć należących do 31 osób, udostępnionym przez California Institute of Technology [1]. 4 zdjęcia zostały odrzucone z powodu braku rozpoznania twarzy w pozycji frontalnej.

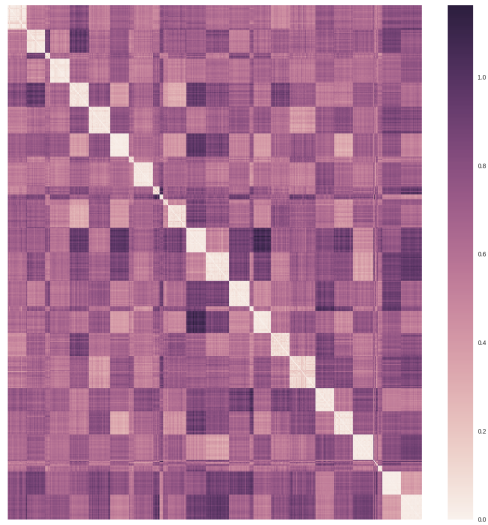
Weryfikację przetestowano poprzez rozważenie wszystkich możliwych par 446 zdjęć i użyciu klasyfikatorów do określenia odległości euklidesowej między tymi zdjęciami. Rysunek 2 ilustruje macierze odległości między poszczególnymi parami zdjęć.

Na podstawie par odległości wyliczono również wskaźniki błędów (FAR i FRR). Przyjmując zmienny próg $t \in [0, 1]$, wektory cech dla dwóch twarzy rozważano jako należące do jednej osoby, jeśli ich wzajemna odległość jest mniejsza lub równa niż t i należące do różnych osób w przeciwnym przypadku. Wartości wskaźników obrazuje wykres 3.

Na podstawie wyznaczonych FAR i FRR można określić, że klasyfikator normalizacyjny osiągnął EER w okolicy 2.5%, zaś klasyfikator oparty na sieci ResNet — zerowy współczynnik równego błędu. W przypadku weryfikacji oba podejścia uzyskały więc dość dobre wyniki na testowanym zbiorze.

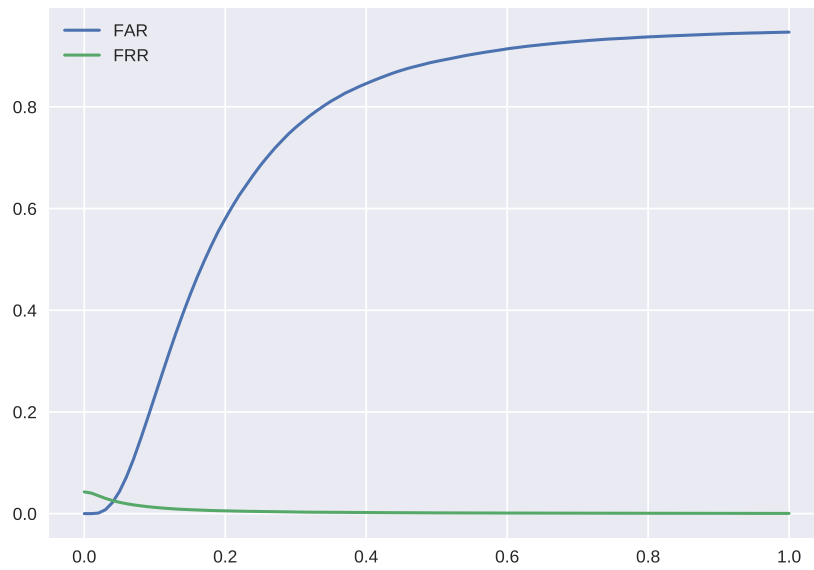


(a) Macierz dla podejścia normalizacyjnego po przetworzeniu odległości wg wzoru $d' = 1 - \frac{1}{d+1}$.

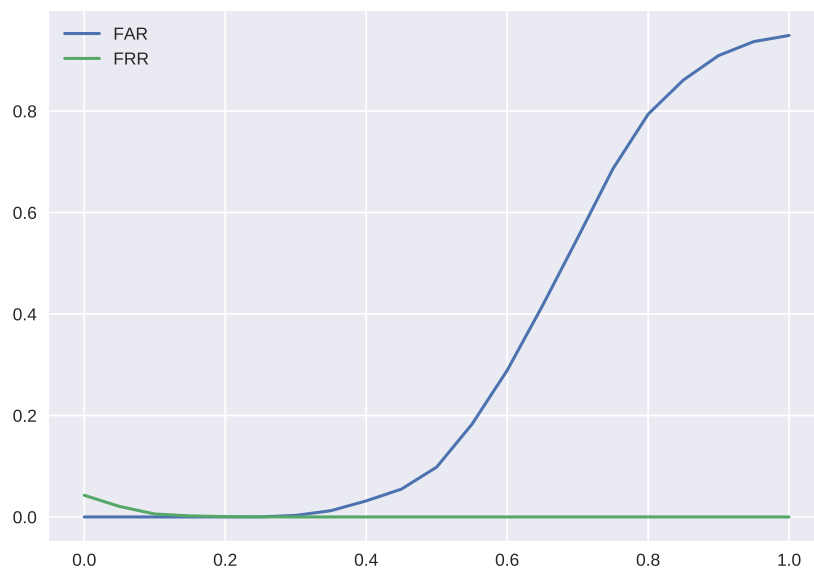


(b) Macierz dla klasyfikatora ResNet.

Rysunek 2: Macierze wzajemnych odległości dla poszczególnych par obrazów ze zbioru testowego. Zauważalne są ciemniejsze bloki wzdłuż przekątnej, obrazujące podobieństwo wielu obrazów przedstawiających jedną osobę.



(a) Wskaźniki FAR i FRR dla podejścia normalizacyjnego.



(b) Wskaźniki FAR i FRR dla klasyfikatora ResNet.

Rysunek 3: Wartości wskaźników FAR i FRR w zależności od przyjętego progu klasyfikacji w zadaniu weryfikacji tożsamości na podstawie par zdjęć.

Test dla zadania identyfikacji przeprowadzono dla każdego obrazu poprzez obliczanie wektorów cech dla wszystkich twarzy i wyborze najbliższego do wektora wybranego obrazu (w sensie normy euklidesowej) wektora z reszty zbioru. Taka metoda eliminuje możliwość uzyskania zerowej odległości rozważanego obrazu od tego samego zbioru w zbiorze porównawczym.

Wyniki testu obrazują macierze pomyłek przedstawione na rysunku 4.

Oba klasyfikatory popełniły niski odsetek pomyłek. Klasyfikator normalizacyjny okazał się jednak dużo gorszy. Przyglądając się przypadkom, w których zaszła pomyłka, określono, że jest to spowodowane brakiem uwzględniania kolorów obrazu przy klasyfikacji — zdarzyły się przypadki, w których dwie twarze (jedna z zarostem, jedna bez) zostały uznane jako należące do jednej osoby z powodu zbliżonej geometrii. Z kolei klasyfikator ResNet, który stosuje warstwy konwolucyjne, unika tego typu błędów i wykazuje znacznie mniejszą liczbę pomyłek (zdarzyły się tylko 4 razy).

5 Podsumowanie

Rozpoznawanie człowieka na podstawie zdjęć twarzy stanowi przyszłościowy aspekt badań biometrycznych. Jak widać po uzyskanych wynikach, istniejące podejścia oparte na uczeniu maszynowym i sieciach neuronowych uzyskują bardzo dobre wyniki w zadaniu weryfikacji i identyfikacji człowieka. Dzięki rosnącej mocy obliczeniowej nowoczesnych urządzeń, rozwiązania takie przenoszą się również na platformy mobilne, dając perspektywę globalnej adaptacji metod opartych na twarzy na potrzeby identyfikacji i weryfikacji tożsamości.

Literatura

- [1] Computational Vision Group, „Archive”, California Institute of Technology. [Online]
Dostępne: <http://www.vision.caltech.edu/html-files/archive.html>. [Dostęp 12 maja 2019]
- [2] da Costa-Luis, C. i inni, „tqdm”. [Online]
Dostępne: <https://tqdm.github.io/>. [Dostęp 11 maja 2019]
- [3] Dalal, N., Triggs, B., „Histograms of Oriented Gradients for Human Detection”, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- [4] He, K., Zhang, X., Ren, S., Sun, J., „Deep Residual Learning for Image Recognition”, [arXiv:1512.03385](https://arxiv.org/abs/1512.03385), 2015. [Online]
Dostępne: <https://arxiv.org/abs/1512.03385>. [Dostęp 11 maja 2019]
- [5] „Matplotlib: A 2D graphics environment”, *Computing In Science & Engineering*, tom 9, nr 3, s. 90–95, 2007.
- [6] King, D. i inni, „dlib C++ Library”. [Online]
Dostępne: <http://dlib.net/>. [Dostęp 11 maja 2019]
- [7] King, D., „dlib-models”. [Online]
Dostępne: <https://github.com/davisking/dlib-models>. [Dostęp 11 maja 2019]

- [8] McKinney, W., „Data Structures for Statistical Computing in Python”, *Proceedings of the 9th Python in Science Conference*, s. 51–56, 2010.
- [9] Oliphant, T.E., *A Guide to NumPy*, Trelgol Publishing, Stany Zjednoczone, 2006.
- [10] Sagonas, C., Zafeiriou, S., „Facial point annotations”. [Online]
Dostępne: <https://iug.doc.ic.ac.uk/resources/facial-point-annotations/>.
[Dostęp 11 maja 2019]
- [11] Waskom, M. i inni, „**seaborn**: statistical data visualization”. [Online]
Dostępne: <https://seaborn.pydata.org/>. [Dostęp 11 maja 2019]