

POLITECHNIKA WROCŁAWSKA

WYDZIAŁ ELEKTRONIKI

KIERUNEK: Automatyka i Robotyka (AIR)
SPECJALNOŚĆ: Technologie informacyjne
w systemach automatyki (ART)

PRACA DYPLOMOWA INŻYNIERSKA

Biometryczny system kontroli dostępu -
przetwarzanie obrazów i przygotowanie danych
dla sieci neuronowych

Biometric access control system - image and
data preprocessing for neural nets

AUTOR:
Jędrzej Kozal

PROWADZĄCY PRACĘ:
dr inż. Piotr Ciskowski

OCENA PRACY:

Rozdział 1

Wstęp

Celem pracy jest porównanie metod i algorytmów umożliwiających wykrycie twarzy na zdjęciu, oraz wyznaczenie wektora uczącego dla sieci neuronowej na podstawie wyznaczonego zdjęcia twarzy. Projekt ten jest częścią systemu biometrycznej kontroli dostępu. Za implementację sieci neuronowych oraz analizę algorytmów związanych z sieciami odpowiadał Filip Guzy. Za architekturę systemu, oraz komunikację pomiędzy komponentami odpowiadał Michał Leś.

Celem całego projektu było skonstruowanie prostego systemu kontroli dostępu opartego na rozpoznawaniu twarzy. System ma określać czy na wykonanym zdjęciu występuje twarz. Ponadto przedmiotem pracy było wyznaczenie sposobu na ekstrakcję cech ze zdjęcia i przygotowanie wektora uczącego dla sieci. Zadaniem sieci neuronowej jest klasyfikacja czy danemu użytkownikowi przyznano wcześniej dostęp czy też nie.

System oryginalnie został zaprojektowany na mikrokomputer Rasbery Pi. W projekcie wykorzystano Rasbery Pi 3. Przy projektowaniu systemu fakt ten miał niebagatelne znaczenie, ze względu na ograniczone pokłady mocy obliczeniowej. Architektura systemu została zaprojektowana w taki sposób, aby system można było z łatwością przenosić na inne platformy.

1.1 Omówienie zagadnienia

1.2 Omawiany komponent jako część większego systemu

Jak wspomniano we wstępie komponent opisany w niniejszej pracy jest częścią większego systemu. W tej sekcji pracy przedstawiono założenia projektowe, zaprezentowano szerszy ogólny obraz na cały system. Omówiono komponent z programistycznego punktu widzenia: omówiono wzorce projektowe, praktyki jakich przestrzegano projektując komponent oraz uzasadniono wybór narzędzi.

1.2.1 Wykorzystane wzorce projektowe

W celu ułatwienia pracy współpracownikom w projekcie został utworzony wydzielony łatwo wymienialny komponent. W celu ułatwienia pracy badawczej komponent został podzielony na mniejsze elementy z wykorzystaniem obiektowych wzorców projektowych. Do reprezentacji całego komponentu wybrano wzorec fabryka. Zapewniono w ten sposób elastyczność użytkownikom klasy, ponieważ mogą oni zdecydować w jaki sposób współrzędne będące wynikiem działania algorytmów mają być przedstawione. Zapewniono klasę bazową do przechowywania współrzędnych, po której dziedziczą poszczególne reprezentacje, takie jak bezpośrednie przechowywanie danych w pamięci programu, czy zapis na pamięć nieulotną w postaci pliku z rozszerzeniem .npy z biblioteki numpy. W celu zwiększenia elastyczności w obrębie komponentu do wyboru algorytmów dwukrotnie wykorzystano wzorec strategia. Stworzono klasy bazowe do reprezentacji podstawowych własności algorytmu, które następnie implementowano w klasach pochodnych. Modyfikacja ta ułatwiła pracę badawczą i usystematyzowała strukturę projektu.

1.2.2 Wykorzystane zasady i dobre praktyki programowania

W trakcie realizacji projektu oprócz wykorzystania wzorców projektowych posługiwano się także dobrymi zasadami SOLID oraz clean code. Pozwalają one na zachowanie większego porządku oraz czytelności kodu. Dobra organizacja kodu oraz porządek ułatwiają rozwijanie projektu oraz systematyzują pracę. Nie zdecydowano się na wykorzystanie testów jednostkowych ze względu na dynamiczny charakter projektu. Definiowanie zachowań w testach jednostkowych jest kosztowne, a częste zmiany powodują że praca ta momentami byłaby zbędna.

1.2.3 Wykorzystane biblioteki, narzędzia i zasoby

Projekt systemu biometrycznej kontroli dostępu został zrealizowany całkowicie w języku programowania Python. Decyzja ta była podyktowana głównie znaczącą ilością gotowych bibliotek, które znacząco ułatwiają pracę. Wykorzystana została wersja języka 2.7.9, ponieważ biblioteki języka Python często wymagają wersji 2.7 lub wyższej. W omawianym komponencie wykorzystano biblioteki NumPy, SciPy, scikit-learn, openCV oraz Dlib. NumPy to biblioteka zawierająca wiele algorytmów do realizacji obliczeń numerycznych oraz implementację wielu matematycznych narzędzi związanych z algebrą liniową. Implementacja algorytmów macierzowych z NumPy cieszy się dużą popularnością, dlatego warto poświęcić czas na poznanie tej biblioteki, ponieważ stanowi ona fundament wielu innych projektów i bibliotek. SciPy jest biblioteką wykorzystywaną do obliczeń naukowych i inżynierskich. Zawiera algorytmy umożliwiające przeprowadzanie obliczeń w wielu dziedzinach, dlatego również jest wykorzystywana jako baza innych projektów. Instalacja SciPy jest prerekwizytem do instalacji scikit-learn. Scikit-learn to biblioteka zawierająca wiele algorytmów z dziedziny machine learning oraz pattern recognition. Jest to kluczowa biblioteka w omawianym projekcie, ponieważ dostarcza najistotniejszych narzędzi. Jej zastosowanie znacznie ułatwiło pracę. OpenCv stanowi potężny zbiór narzędzi do analizy oraz przetwarzania obrazów. Rozpowszechniana w ramach licencji . Twórcy biblioteki skupiają się na działaniu w czasie rzeczywistym i zapewniają wielordzeniowe przetwarzanie. Jest to znaczące ułatwienie biorąc pod uwagę podstawową platformę sprzętową, na której był realizowany projekt. Dlib jest biblioteką przewidzianą głównie dla języka C++,

jednakże możliwe jest także wykorzystanie jest w Pythonie. Zawiera głównie algorytmu z dziedziny ML. W omawianym projekcie została wykorzystana do zbadania możliwości alternatywnych metod detekcji twarzy.

Wszystkie wymienione biblioteki są rozpowszechniane w ramach licencji new-BSD lub 3-clause BSD, co oznacza że wolno je wykorzystywać w celach akademickich lub komercyjnych. Zastosowanie bibliotek dostarczających zoptymalizowanych algorytmów znacznie ułatwiło i usystematyzowało pracę nad projektem.

Na początku pracy nad projektem podjęto próbę implementacji własnej wersji algorytmu PCA, aby lepiej zrozumieć i dokładnie prześledzić jego działanie. Algorytm ten działał poprawnie, ale względu na dużą złożoność obliczeniową zdecydowano się na korzystanie z wersji udostępnionej w bibliotece scikit-learn. W trakcie późniejszych prac nad systemem znaleziono informację na temat tego co powodowało tak wysoką złożoność obliczeniową oraz sposób na jej uniknięcie. Zagadnienie to zostało opisane w rozdziale Realizacja algorytmu.

Diagramy i rysunki wykonano z wykorzystaniem strony draw.io. Do wszystkich zapożyczonych obrazów zostały podane źródła.

Do przeprowadzenia badań jako bazy twarzy wykorzystano zbiór zdjęć dostępny w ramach biblioteki scikit-learn The Olivetti faces dataset. Posłużył on jako baza do wyznaczenia wartości własnych przestrzeni zdjęć twarzy.

Rozdział 2

Implementacja algorytmów – detekcja twarzy

Pierwszym wymaganiem po zdolności do akwizycji obrazów jest określenie czy uzyskane zdjęcie przedstawia twarz. W niniejszym rozdziale omówiono możliwe metody detekcji twarzy na dwuwymiarowym obrazie. Szczegółowo omówiono przyjęte rozwiązanie, oraz podano uzasadnienie dlaczego zostało ono wybrane.

2.1 Kaskady Haara

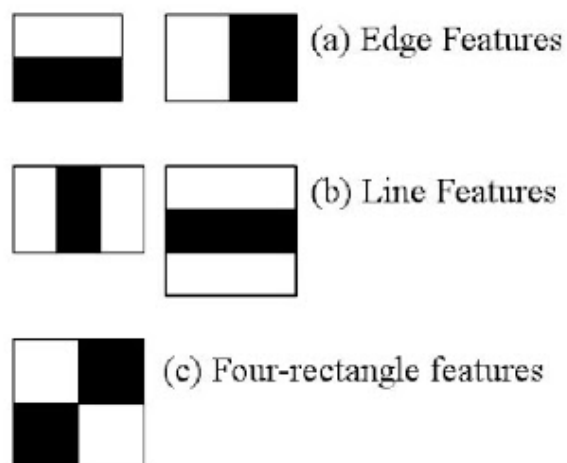
Algorytm Kaskad Haara (lub detektor Viola-Jones) wykorzystuje serię cech do określenia czy na zdjęciu w określonym miejscu znajduje się szukany obiekt. Cecha jest tutaj rozumiana jako prostokąt podzielony na części. Pod każdą z części dodawane są wartości pikseli, następnie wartości z obu obszarów są od siebie odejmowane. Pomysł ten przypomina falkę Haara, stąd nazwa algorytmu. Tak uzyskana wartość musi przekroczyć odpowiedni próg, który decyduje o tym czy w danym miejscu rozpoznano twarz. Jak podano w [5], aby móc szybciej obliczać wartości pikseli pod odpowiednim obszarem wykorzystuje się obraz kumulacyjny (ang. integral image). Jest on zdefiniowany jako:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (2.1)$$

Funkcja ta przypomina dyskretną dystrybuantę dwuwymiarowej zmiennej losowej. Jej zastosowanie pozwala na obliczenie sumy jasności pod dowolnie wielkim obszarem, w dowolnym miejscu obrazu w stałym czasie.

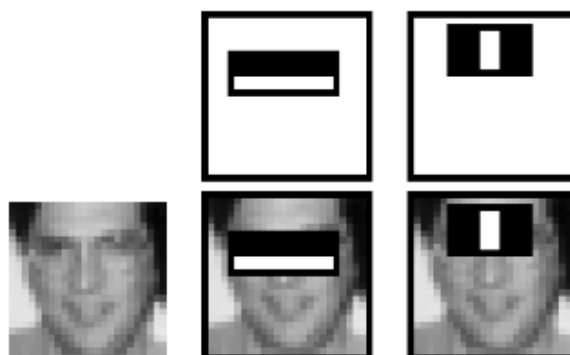
Nawet dla najmniejszych obrazów można rozważać wiele cech. W celu ograniczenia ilości analizowanych cech wykorzystano algorytm AdaBoost [4]. Algorytm ten dysponując zbiorem uczącym oraz zbiorem słabych hipotez przydziela wagi hipotezom, oceniając ich skuteczność. Dodatkowo wagi otrzymują także przykłady ze zbioru uczącego. Każdorazowo kiedy dany przykład zostanie źle zaklasyfikowany, jego waga jest zwiększana.

Wykorzystanie AdaBoost pozwoliło na wybranie cech generujących najmniejszy błąd. Ilość użytych cech wpływa na jakość algorytmu, jednocześnie zwiększając czas potrzebny na obliczenia. Jak można zauważyć na 2.2 najlepsze cechy mają dość intuicyjną inter-



Rysunek 2.1: Różne rodzaje cech. Od sumy wartości pikseli po czarnym obszarze jest odejmowana wartość sumy pikseli pod białym polem.

Źródło: https://docs.opencv.org3.3.0d7d8btutorial_py_face_detection.html



Rysunek 2.2: Dwie cechy generujące najmniejszy błąd z wykorzystaniem AdaBoost.

Źródło: https://docs.opencv.org3.3.0d7d8btutorial_py_face_detection.html



Rysunek 2.3: Wykrywanie punktów charakterystycznych twarzy.

W lewej części oryginalny obraz z kamertki, wprawiej części obraz z zaznaczonymi rezultatami detekcji. Wykrozystano 68 cechy.

pretację: spodziewamy się że obraz w okolicach oczu będzie ciemniejszy niż w okolicach policzków, oraz że środek nosa będzie jaśniejszy niż obszary po bokach.

Dysponując zbiorem cech można przystąpić do klasyfikacji obrazu. Odbywa się to kaskadowo. W pierwszym etapie wykorzystuje się klasyfikator złożony najczęściej z jednej lub dwóch cech. Klasyfikacja jest przeprowadzana na całym obrazie, aby wykryć gdzie może znajdować się twarz. W kolejnych etapach ilość cech jest zwiększana (najczęściej wykładniczo). Jeśli w danym obszarze nie została wykryta twarz zostaje on pominięty w dalszych etapach. Pozwala to znacząco ograniczyć ilość wykonywanych operacji. Zaleca się aby w pierwszym etapie wykorzystywać cechę z wysokim współczynnikiem fałszywych negatywnych rezultatów (ang. false negative) - w tym przypadku obszarów zawierających twarz błędnie zaklasyfikowanych jako nie będących twarzą.

Kaskady Haara są szybkie, przez co chętnie wykorzystuje się je w zastosowaniach wymagającego przetwarzania obrazów w czasie rzeczywistym. Było to także motywacją przy wyborze tej metody w omawianym systemie. Rozwiązanie daje to zadowalające wyniki przy jednoczesnym szybkim działaniu. W rozważanym systemie kosztowność obliczeniowa musi być brana pod uwagę, biorąc pod uwagę, że oryginalna platforma sprzętowa na którą został przewidziany system to Rasbery Pi.

2.2 Inne sposoby rozwiązania problemu

Do całego problemu można zastosować inne podejście polegające na szukaniu charakterystycznych punktów twarzy. W pracy [6] wykorzystano boosting drzewa gradientu do stworzenia kaskady regresorów. Dysponując zbiorem uczącym zawierającym zdjęcia twarzy oraz współrzędne punktów charakterystycznych można wyznaczyć kaskadę regresorów liniowych. Zakładamy że estymowany kształt punktów charakterystycznych dowolnej twarzy może być wyrażony jako kombinacja liniowa kształtu punktów charakterystycznych z odpowiednio dużego zbioru uczącego. Każdy regresor to drzewo decyzyjne, które w liściach przechowuje wektor, który powinien zostać dodany do aktualnej estymacji punktów charakterystycznych (zbiór współrzędnych x i y które powinny być kolejno dodane do każdej współrzędnej). W pierwszym kroku jako estymacja jest wykorzystywany średni kształt z bazy zdjęć uczących. Następnie każdy regresor delikatnie poprawia estymację, przesuwając punkty w odpowiednią stronę. Na rys. 2.3 przedstawiono efekty detekcji twarzy z

wykorzystaniem implementacji z biblioteki Dlib.

Omawiane rozwiązanie umożliwia całkowicie inne podejście do problemu ekstrakcji cech. Mając estymatę charakterystyczne punkty twarzy, oraz mierząc odległości między poszczególnymi punktami można przygotować wektor uczący mający bardzo fizyczne interpretację. Dodając informację na temat koloru oczu czy skóry, można skonstruować zupełnie inne wektory cech niż proponowany w tej pracy.

Rozdział 3

Implementacja algorytmów – ekstrakcja cech

Niniejszy rozdział traktuje o algorytmach umożliwiających ekstrakcję cech twarzy ze zdjęć. Dysponując zdjęciem twarzy należy zaproponować sposób utworzenia wektora, będącego reprezentatywnym odzwierciedleniem ilościowych cech ze zdjęcia i jednocześnie umożliwi sieci neuronowej dokonanie klasyfikacji. Zadanie to jest bardzo abstrakcyjne, zależy od wielu czynników i trudno jest określić rezultaty.

W rozdziale tym skupiono się głównie nad podejściem polegającym na zastosowaniu algorytmów redukcji wymiarowości, takich jak PCA czy SVD. Jest to dość popularne rozwiązanie i w przypadku PCA można uzyskać wyniki mające wizualną interpretację, co jest bardzo pomocne przy pracy nad tego typu problemem.

Obraz po wykryciu twarzy jest obcinany do kwadratu, na którym wykryto twarz, następnie obraz jest obcinany do rozmiaru 64×64 , w celu ujednolicenia rozmiarów przetwarzanych obrazów oraz zwiększenia szybkości przetwarzania.

3.1 Principle Components Analysis

Principle Components Analysis (PCA) jest algorytmem redukcji wymiarowości danych. Mając określony zbiór obserwacji, będący reprezentatywną reprezentacją próbą analizowanej przestrzeni należy wyznaczyć zbiór wektorów bazowych, tak aby wariancja danych z zbioru obserwacji względem nowo wyznaczonych wektorów bazowych była jak najwyższa.

3.1.1 Wstęp matematyczny

Przedstawione tutaj wyprowadzenie jest podawane za Pattern recognition and machine learning.

Zakładamy że dany jest zbiór D -wymiarowych wektorów obserwacji: $\mathbf{x}_n, n \in 1, \dots, N$. Celem jest znalezienie nowej M -wymiarowej bazy ($M < D$), takiej że wariancja zbioru wektorów \mathbf{x}_n po rzutowaniu na nową bazę jest największa. W celu pokazania procesu i rozumowania przyjmijmy że szukamy jednego wektora \mathbf{u}_1 . Wektor ten, ponieważ ma

stanowiąc bazę musi być D-wymiarowy. Dodatkowo przyjmujemy że wektor \mathbf{u}_1 jest jednostkowy, co zapiszemy w postaci warunku:

$$\mathbf{u}_1^T \mathbf{u}_1 = 1 \quad (3.1)$$

Rzut wektora \mathbf{x} na kierunek wyznaczany przez wektor \mathbf{u} jest dany przez: $u^T x$. Fakt ten ma proste uzasadnienie geometryczne:

$$\begin{aligned} \mathbf{u}^T \mathbf{x} &= \mathbf{u} \circ \mathbf{x} \\ &= \|\mathbf{u}\| \|\mathbf{x}\| \cos \angle(\mathbf{u}, \mathbf{x}) \\ &= \|\mathbf{x}\| \cos \angle(\mathbf{u}, \mathbf{x}) \end{aligned} \quad (3.1)$$

Przyjmimy średnią jako estymator wartości oczekiwanej:

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \quad (3.2)$$

Wariancja zmiennej losowej X jest dana przez:

$$\begin{aligned} \text{var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \end{aligned}$$

Wariancja jest miarą rozrzutu wartości zmiennej losowej. Im większe różnice między wartościami przyjmowanymi przez zmienną losową, tym większa jest wariancja.

Estymator wariancji dla próby losowej x_n :

$$s^2 = \frac{1}{N} \sum_{n=1}^N (x_n - \bar{x})^2$$

Kowariancja między dwoma zmiennymi losowymi jest dana wzorem:

$$\begin{aligned} \text{cov}[X, Y] &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned} \quad (3.3)$$

Wzór ten można uogólnić dla wektorów losowych:

$$\begin{aligned} \text{cov}[\mathbf{X}, \mathbf{Y}] &= \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{Y}^T - \mathbb{E}[\mathbf{Y}^T])] \\ &= \mathbb{E}[\mathbf{X}\mathbf{Y}^T] - \mathbb{E}[\mathbf{X}]\mathbb{E}[\mathbf{Y}^T] \end{aligned}$$

Definiujemy macierz kowariancji S dla zbioru obserwacji \mathbf{x}_n :

$$S = \frac{1}{N} \sum_{n=1}^N (\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T \quad (3.4)$$

Jak łatwo zauważyć w powyższym wzorze wartości oczekiwane zostały zastąpione średnimi, ponieważ nie dysponujemy rozkładami zmiennych losowych, tylko zbiorem obserwacji.

Definiujemy wariancję zbioru obserwacji względem wektora u_1 :

$$s^2 = \frac{1}{N} \sum_{n=1}^N (\mathbf{u}_1^T \mathbf{x}_n - \mathbf{u}_1^T \bar{\mathbf{x}})^2 = \mathbf{u}_1^T S \mathbf{u}_1 \quad (3.5)$$

Chcąc maksymalizować wariancję s^2 musimy pamiętać o warunku 3.1. Otrzymujemy więc problem optymalizacji polegający na znalezieniu maksimum warunkowego funkcji $f(\mathbf{u}_1)$ z ograniczeniem $g(\mathbf{u}_1)$:

$$\begin{aligned} f(\mathbf{u}_1) &= \mathbf{u}_1^T S \mathbf{u}_1 \\ g(\mathbf{u}_1) &= 1 - \mathbf{u}_1^T \mathbf{u}_1 = 0 \end{aligned}$$

Jest to problem, który można łatwo rozwiązać z wykorzystaniem mnożników Lagrange'a. W analizowanym przypadku Lagrangian przyjmuje postać:

$$\begin{aligned} \mathcal{L}(\mathbf{u}_1, \lambda) &= f(\mathbf{u}_1) - \lambda g(\mathbf{u}_1) \\ &= \mathbf{u}_1^T S \mathbf{u}_1 - \lambda(1 - \mathbf{u}_1^T \mathbf{u}_1) \end{aligned}$$

Różniczkując względem \mathbf{u}_1 oraz λ oraz przyrównując do zera uzyskujemy układ równań:

$$\begin{cases} \nabla_{\mathbf{u}_1} \mathcal{L}(\mathbf{u}_1, \lambda) = \nabla_{\mathbf{u}_1} \mathbf{u}_1^T S \mathbf{u}_1 - \nabla_{\mathbf{u}_1} \lambda \mathbf{u}_1^T \mathbf{u}_1 = 0 \\ \frac{\partial}{\partial \lambda} \mathcal{L}(\mathbf{u}_1, \lambda) = 1 - \mathbf{u}_1^T \mathbf{u}_1 = 0 \end{cases} \quad (3.6)$$

Zajmijmy się pierwszym równaniem z (6). Odjemnik można bardzo łatwo uprościć:

$$\nabla_{\mathbf{u}_1} \lambda \mathbf{u}_1^T \mathbf{u}_1 = \lambda \nabla_{\mathbf{u}_1} (u_{1_1}^2 + u_{1_2}^2 + \dots + u_{1_D}^2) = 2\lambda \mathbf{u}_1 \quad (3.7)$$

Odjemna wymaga więcej przekształceń:

$$\nabla_{\mathbf{u}_1} \mathbf{u}_1^T S \mathbf{u}_1 = \nabla_{\mathbf{u}_1} \begin{pmatrix} u_{1_1} & u_{1_2} & \dots & u_{1_D} \end{pmatrix} \begin{pmatrix} s_{1_1} u_{1_1} + s_{1_2} u_{1_2} + \dots + s_{1_D} u_{1_D} \\ s_{2_1} u_{1_1} + s_{2_2} u_{1_2} + \dots + s_{2_D} u_{1_D} \\ \vdots \\ s_{D_1} u_{1_1} + s_{D_2} u_{1_2} + \dots + s_{D_D} u_{1_D} \end{pmatrix} \quad (3.8)$$

$$\begin{aligned} &= \nabla_{\mathbf{u}_1} [u_{1_1}(s_{1_1} u_{1_1} + s_{1_2} u_{1_2} + \dots + s_{1_D} u_{1_D}) \\ &+ u_{1_2}(s_{2_1} u_{1_1} + s_{2_2} u_{1_2} + \dots + s_{2_D} u_{1_D}) \\ &+ \dots + u_{1_D}(s_{D_1} u_{1_1} + s_{D_2} u_{1_2} + \dots + s_{D_D} u_{1_D})] \end{aligned} \quad (3.9)$$

$$= \begin{pmatrix} 2s_{1_1} u_{1_1} + 2s_{1_2} u_{1_2} + \dots + 2s_{1_D} u_{1_D} \\ 2s_{2_1} u_{1_1} + 2s_{2_2} u_{1_2} + \dots + 2s_{2_D} u_{1_D} \\ \vdots \\ 2s_{D_1} u_{1_1} + 2s_{D_2} u_{1_2} + \dots + 2s_{D_D} u_{1_D} \end{pmatrix} \quad (3.10)$$

$$= 2S \mathbf{u}_1 \quad (3.11)$$

W przejściu między (9) a (10) korzystamy z faktu że macierz S (macierz kowariancji) jest symetryczna.

Wracając do (6):

$$\begin{aligned} 2S \mathbf{u}_1 - 2\lambda \mathbf{u}_1 &= 0 \\ S \mathbf{u}_1 &= \lambda \mathbf{u}_1 \end{aligned} \quad (3.12)$$

Jest równanie na wartości własne macierzy S , więc \mathbf{u}_1 musi być wektorem własnym macierzy S , λ wartością własną macierzy S .

Mnożąc (11) lewostronnie przez \mathbf{u}_1^T otrzymujemy:

$$\mathbf{u}_1^T S \mathbf{u}_1 = \lambda \quad (3.13)$$

Lewa część (13) to wariancja, która ma być maksymalizowana. Aby uzyskać jak największą wariancję zbioru obserwacji po zrzutowaniu na kierunek wyznaczany przez \mathbf{u}_1 należy wybrać największą wartość własną macierzy S . Odpowiadający jej wektor własny jest szukanym wektorem \mathbf{u}_1 . Dla kolejnych wektorów \mathbf{u} należy znaleźć kolejne największe wartości własne i wektory własne.

3.1.2 Realizacja algorytmu

Korzystając z PCA można zmniejszyć wymiarowość danych w taki sposób aby zachować najwięcej informacji. Działanie algorytmu można zasadniczo podzielić na dwa etapy. Pierwszy to uzyskanie bazy wektorów, która zostanie wykorzystana w drugim etapie do uzyskania nowej reprezentacji wektorów z oryginalnej przestrzeni.

W pierwszym etapie należy obliczyć wartości własne macierzy kowariancji, jak to zostało pokazane w poprzednim dziale. Przyjmijmy że dysponujemy macierzą pomiarów:

$$X_{D \times N} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_N \end{pmatrix} = \begin{pmatrix} x_{1_1} & x_{1_2} & \dots & x_{1_N} \\ x_{2_1} & x_{2_2} & \dots & x_{2_N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{D_1} & x_{D_2} & \dots & x_{D_N} \end{pmatrix} \quad (3.14)$$

Zdefiniujmy macierz A , powstałą w wyniku odjęcia od każdego wiersza średniej:

$$A_{D \times N} = \begin{pmatrix} \mathbf{x}_1 - \bar{\mathbf{x}} & \mathbf{x}_2 - \bar{\mathbf{x}} & \dots & \mathbf{x}_N - \bar{\mathbf{x}} \end{pmatrix} \quad (3.15)$$

Dysponując macierzą A , można zapisać S w postaci:

$$S_{D \times D} = \frac{1}{N} A A^T \quad (3.16)$$

Co jest całkowicie zgodnie z definicją 3.3. Kolejnym krokiem będzie wyznaczenie wektorów i wartości własnych, posortowanie wektorów własnych według nierosnącej wartości własnych oraz wybranie M wektorów własnych.

Należy zwrócić uwagę na wymiar macierzy S . Przypomnijmy że D to wymiar wektora pomiarowego, co w przypadku zdjęć oznacza wartość każdego piksela przeniesiona kolejno do elementów wektora. W pracy tej zdjęcia przed podaniem na wejście algorytmu ekstrakcji cech były konwertowane do skali szarości i obcinane do rozmiaru 64×64 pikseli, co oznacza że wektor pomiarowy zawierał 4096 współrzędnych. Biorąc pod uwagę złożoność obliczeniową algorytmów szukania wektorów własnych może to powodować problemy z wydajnością. Można temu zapobiec poprzez zamianę miejscami macierze A i A^T w iloczynie. Obliczając wartości własne macierzy $A^T A$ musimy wrócić do D -wymiarowości przestrzeni, co jest możliwe z wykorzystaniem następującej własności:

$$u_i = A v_i \quad (3.17)$$

Gdzie v_i oznacza wartość własną macierzy $A^T A$. Poprzez zamianę miejscami macierzy A i A^T zyskujemy inny wymiar macierzy. Iloczyn ma teraz rozmiar $N \times N$. Przypomnijmy że N to liczność zbioru obserwacji. Zazwyczaj N jest znacznie niższe od D , co może przyspieszyć obliczenia. Korzystając z tego uproszczenia należy pamiętać że zostaje nałożone ograniczenie co wielkości M . Jako że zostają obliczone wektory własne macierzy o rozmiarze $N \times N$, może ich być co najwyżej N . Oznacza że ilość wymiarów danych po redukcji nie

może przekraczać ilość danych ze zbioru obserwacji ($M < N$). Ograniczenie to nie jest zbyt dotkliwe, gdy dysponujemy odpowiednio wielkim zbiorem testowym. W przypadku tej pracy dostępność danych stanowiących podstawę nie stanowiła problemu.

Po konwersji do oryginalnej wymiarowości dysponujemy bazą M wektorów D -wymiarowych. Zbiór M wektorów jest nazywany wartościami własnymi i zostanie wykorzystany w drugim etapie. Dla uproszczenia wektory własne zostaną włączone do jednej macierzy:

$$U = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_M \end{pmatrix} \quad (3.18)$$

W pewnym sensie wyznacza on reguły przy pomocy których konwertujemy wektory D -wymiarowe na M -wymiarowe.

Drugim etapem działania algorytmu jest redukcja wymiarowości wektora \mathbf{w} spoza zbioru obserwacji. Obrazowo można powiedzieć że na podstawie nowego zdjęcia, niewykorzystanego w poprzednim etapie chcemy uzyskać reprezentujący je wektor liczb, w przestrzeni o niższym wymiarze. Odbywa się to zasadniczo w dwóch bardzo prostych krokach: Pierwszym jest odjęcie od nowego wektora średniej. Drugim jest rzutowanie różnicy $\mathbf{w} - \bar{\mathbf{x}}$ na wcześniej uzyskane wartości własne. Uzyskanie współrzędnych w M -wymiarowej przestrzeni wektora w można więc zapisać jako:

$$U(\mathbf{w} - \bar{\mathbf{x}}) \quad (3.19)$$

Jak łatwo można zauważyć $\bar{\mathbf{x}}$ oraz U zależą bezpośrednio od zbioru obserwacji jakim dysponujemy na początku. Dobranie zbioru reprezentatywnego dla danej przestrzeni jest kluczowe. Pierwszy etap działania algorytmu polegający na generacji wektorów własnych definiuje przekształcenie. Drugi etap działania pozwala na redukcję wymiarowości dowolnego D -wymiarowego wektora zgodnie z regułami ustalonymi w pierwszym kroku.

3.1.3 Interpretacja

Przyjrzyjmy się przez moment co tak na prawdę zyskujemy w wyniku działania algorytmu PCA. Przez odjęcie średniej przesuwamy środek nowo powstałego układu współrzędnych do punktu gdzie znajdowała się średnia zbioru obserwacji. Dodatkowo po odjęciu średniej wektor reprezentuje jedynie różnicę między średnią (środkiem nowego układu współrzędnych), a punktem końcowym. Jest więc reprezentacją unikatowych cech w zbiorze obserwacji. Wybranie wektorów własnych, będących osiami - kierunkami z największą wariancją powoduje że odrzucając część bazy wektorów tracimy tak mało wektorów jak to jest tylko możliwe. W celach badawczych można wrócić do oryginalnej przestrzeni, aby porównać błąd przybliżenia oraz przeanalizować jak bardzo reprezentacyjne są uzyskane wyniki. W przypadku zdjęć możliwa jest wręcz wizualne porównanie jak bardzo oryginalne zdjęcie i zdjęcie odtworzone po redukcji wymiarowości są podobne do siebie. Rekonstrukcja odbywa się w następujący sposób:

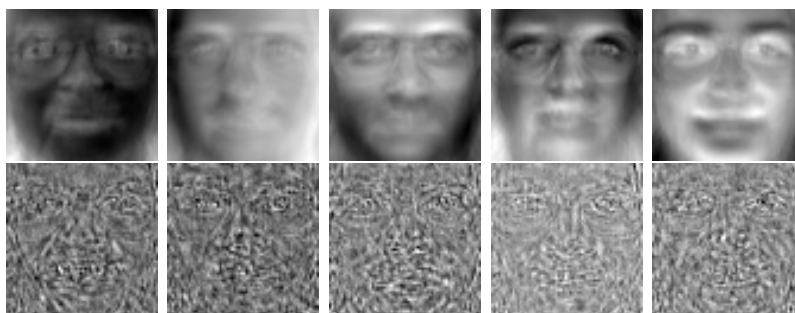
$$v = \bar{x} + \sum_{n=1}^M u_n \quad (3.20)$$

Jak widać na rys. 3.1 wraz ze zwiększeniem się ilości składowych jakość rekonstrukcji zdjęcia się polepsza. Użycie większej ilości składowych nie ma sensu, co można zaobserwować w wyglądzie rekonstrukcji z wykorzystaniem 300 i 400 składowych. PCA jest bardzo



Rysunek 3.1: Porównanie rekonstrukcji twarzy na podstawie różnej liczby wykorzystanych składowych.

Po lewej oryginalne zdjęcie, potem kolejno rekonstrukcja z wykorzystaniem 100, 200, 300 i 400 składowych.



Rysunek 3.2: Porównanie eigenfaces.

Górny wiersz - wektory 0-4, dolny wiersz - wektory 392-396.

lubianą metodą, ponieważ jest bardzo intuicyjna oraz pozwala na wizualne sprawdzenie własnych rezultatów.

W omawianym zagadnieniu wektory własne zyskały specjalną nazwę eigenfaces (twarze własne). Można zobaczyć jak wyglądają aplikując technikę odtwarzania zdjęcia na podstawie współrzędnych z przestrzeni o zredukowanej wymiarowości. Efekty zostały zaprezentowane na rys. 3.2.

Jak można łatwo zaobserwować wektory 0-4 zawierają najbardziej istotne zmiany pomiędzy poszczególnymi obrazami ze zbioru obserwacji. Wektora 392-396 przypominają losowy szum i niosą informację o szczegółach twarzy. Można zauważyć że pominięcie składowych z niższą wartością własną powoduje mniejszą utratę informacji.

3.1.4 PCA jako samodzielny system rozpoznawania twarzy

Principle components analysis powoduje przeniesienie zdjęcia mającego wizualną reprezentację na abstrakcyjny wektor liczb. Umożliwia to skonstruowanie klasyfikatora z wykorzystaniem algorytmów umożliwiających klasyfikację. Jest to zagadnienie dobrze znane w dziedzinie Pattern Recognition (rozpoznawania wzorców) i opracowano wiele rozwiązań tego problemu. Zastosowanie algorytmów typu k-NN (k Nearest Neighbour - k najbliższych sąsiadów) lub NM (Nearest Mean - najbliższy średnia) umożliwia zaklasyfikowanie danego algorytmu do jednej z grup. Jest to możliwe dzięki wykorzystaniu reprezentacji zdjęcia jako wektora liczb. Oba algorytmy dysponują zbiorem uczącym, który opisuje dane grupy do których może zostać zaklasyfikowany dany wzorec. W przypadku NM dla każdej z grup obliczana jest średnia wektorów, następnie obliczana jest odległość między wektorem podlegającym klasyfikacji, a średnimi zbiorów. Wektor zostaje zaklasyfikowany do zbioru,

którego średnia jest w najmniejszej odległości od wektora. W przypadku algorytmu k-NN odległość jest obliczana pomiędzy wszystkimi wektorami ze zbioru uczącego. Następnie zostaje wybrane k najbliższych sąsiadów. Zbiór, którego najwięcej przedstawicieli zostało wybranych wygrywa. Dla przypadku dychotomii jeżeli k jest nieparzyste możliwa jest jednoznaczna klasyfikacja. Gdy ilość zbiorów do których można zaklasyfikować obraz jest większa od dwóch możliwe jest wystąpienie wypadku w którym nie można jednoznacznie zaklasyfikować obiekt. W przypadku systemu kontroli dostępu może być rozsądne wprowadzenie maksymalnej odległości między średnią lub najbliższym sąsiadem a klasyfikowanym wektorem przy której może nastąpić klasyfikacja. Powyżej tej odległości obraz nie powinien być zaklasyfikowany do którejkolwiek z grup. Dodatkowo warto zwrócić uwagę na brak kontroli PCA nad przetwarzanymi danymi. Jest to metoda statystyczna, na jej wejście można podać zdjęcie myszy lub samochodu, które zostanie przetworzone. Nie wiadomo jakie współrzędne można uzyskać, dzięki takiej operacji. System rozpoznawania twarzy powinien więc odrzucić takie zdjęcie na etapie wykrywania pozycji twarzy.

3.2 Inne podejścia do problemu redukcji wymiarowości

Istnieje wiele algorytmów umożliwiających redukcję wymiarowości. Ze względu na rosnącą ilość danych do przetwarzania, algorytmy te zyskały na znaczeniu. W niniejszym fragmencie pracy zostaną omówione zasady działania trzech algorytmów umożliwiających redukcję wymiarowości.

3.2.1 Nonnegative Matrix Factorization

Nonnegative Matrix Factorization (NMF) jest sposobem rozkładu macierzy na dwie mniejsze. Przyjmimy że dana jest macierz $V_{m \times n}$, która ma zostać podana redukcji wymiarowości. Macierz ta zostaje wyrażona jako iloczyn dwóch mniejszych macierzy:

$$V \approx WH \quad (3.21)$$

Zakłada się rozmiary macierzy: $W_{m \times k}$ oraz $H_{k \times n}$, dla $k \leq \min(m, n)$. Dodatkowo przyjmuje się że zarówno W jak i H są nieujemne. Łatwo można zauważyć że stosując taki zapis i przyjmując stosunkowo małą wartość k można zapisać V z wykorzystaniem mniejszej ilości zmiennych. Dodatkowo korzystając ze spostrzeżenia że wektory kolumnowe V mogą zostać zapisane jako:

$$\mathbf{v}_i = W\mathbf{h}_i \quad (3.22)$$

można przyjąć W jako bazę przestrzeni definiowanej przez V .

Dokładny rozkład we wzorze 3.21 nie musi istnieć dla dowolnego k , dlatego nie może zostać umieszczony znak równości. Zwykle w trakcie szukania W i H przyjmuje się że ich iloczyn ma się znajdować w odpowiednio niewielkiej odległości od V . Najczęściej rozważane wartości błędu to:

$$\|A - B\|^2 = \sum_{ij} (A_{ij} - B_{ij})^2$$

$$D(A||B) = \sum_{ij} (A_{ij} \log \frac{A_{ij}}{B_{ij}} - A_{ij} + B_{ij})$$

Pierwsza z prezentowanych norm to znana norma L2, czyli odległość Euklidesowa. Drugie równanie to zmodyfikowana wersja dywergencji Kullbacka-Leiblera dla nieujemnych macierzy. Jest to wielkość informująca nas o tym jak wiele informacji tracimy korzystając z estymacji B rozkładu A . Wielkość ta nie jest normą, ponieważ nie jest symetryczna.

W pracy [7] zaproponowano iteracyjne metody znajdowania W i H dla obu miar błędu. Polegały one na stopniowym poprawianiu wartości obu macierzy. Pierwszy rodzaj polegał na przemnażaniu aktualnej wartości W i H przez odpowiednie współczynniki. Drugi rodzaj opierał się na dodawaniu odpowiedniej poprawki (rozwiązanie przypominające szukanie optimum funkcji z wykorzystaniem gradientu). W obu przypadkach początkowe wartości były losowe, co powodowało, że rezultaty uzyskiwane przez algorytm mogły być różne dla tej samej macierzy V . Dodatkowym ryzykiem jest utknięcie w minimum lokalnym.

W pracy [8] zaproponowano metodę wyznaczania W i H , która nie inicjalizuje losowo wartości początkowych, tylko korzysta z rozkładu SVD (singular-value decomposition - rozkład według wartości osobliwych), który dla macierzy A o wartościach rzeczywistych zapisujemy jako:

$$A = U\Sigma V^T \quad (3.23)$$

U oraz V są ortogonalne, a Σ jest macierzą diagonalną, na przekątnej której leżą tzw. wartości osobliwe. Wektory kolumnowe U są nazywane lewymi wektorami osobliwymi, wektory V są nazywane prawymi wektorami osobliwymi. Autorzy [8] wykorzystali lewe i prawe wektory osobliwe odpowiadające największym wartościom osobliwym Σ do inicjalizacji W i H . Umożliwiło to poprawienie wyników, oraz uzyskanie algorytmu dającego te same rezultaty dla takiego samego wejścia. Dodatkowo czas działania programu korzystającego z tej metody stał się bardziej deterministyczny.

3.2.2 Independent Components Analysis

Rozważmy następujący problem: dana jest obserwacja x będąca liniową kombinacją wielu sygnałów s o niegaussowskim rozkładzie:

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n \quad (3.24)$$

Powższe równanie opisuje wartość j -tej obserwacji. Zarówno x i s są n -wymiarowymi wektorami. Zapis może zostać uogólniony przez zapis macierzowy:

$$\mathbf{x} = A\mathbf{s} \quad (3.25)$$

$A_{n \times n}$ jest nazywane macierzą mieszania (angl. mixing matrix), \mathbf{s} jest nazywane niezależną składową. Celem jest odzyskanie niezależnych składowych \mathbf{s} na podstawie \mathbf{x} . Zauważmy że: $\mathbf{s} = A^{-1}\mathbf{x} = W\mathbf{x}$. W jest nazywana macierzą rozdzielania (angl. unmixing matrix).

Można sobie wyobrazić że problem ten występuje w kontekście analizy sygnału mowy: dysponujemy sygnałem w którym wiele osób mówi jednocześnie a chcemy uzyskać osobne wypowiedzi poszczególnych osób (problem cocktail party).

W [9] podano rozwiązanie problemu opierające się na centralnym twierdzeniu granicznym:

Założmy że wartość oczekiwana \mathbf{x} i \mathbf{s} jest równa 0, oraz wariancja niezależnych składowych jest równa 1. Możemy to zrobić ponieważ dla rozkładów o niezerowej średniej można odjąć średnią. Ograniczenie nałożone na wariancję \mathbf{s} jest spowodowane przez równanie 3.25, konkretniej wartość \mathbf{s} może zostać obliczona z dokładnością do stałej (\mathbf{s} może zostać przemnożone przez dowolne p , jeśli A zostanie przemnożone przez $\frac{1}{p}$). Stosując to ograniczenie unikamy zbyt wielkich wartości niezależnych składowych. Dodatkowo założymy że wszystkie niezależne składowe \mathbf{s} mają taki sam rozkład i są niezależnymi zmiennymi losowymi. Wprowadźmy \mathbf{y} jako liniową kombinację \mathbf{x} :

$$\mathbf{y} = \mathbf{w}^T \mathbf{x} \quad (3.26)$$

Jeśli \mathbf{w} jest fragmentem macierzy W , to \mathbf{y} musi odpowiadać jednej z niezależnych składowych. Nie możemy bezpośrednio wyznaczyć \mathbf{y} , ponieważ nie znamy A i \mathbf{s} , możemy co najwyżej je estymować. Przyjmijmy:

$$\mathbf{z} = A^T \mathbf{w} \quad (3.27)$$

Wykorzystując nową zmienną możemy zapisać:

$$\mathbf{y} = \mathbf{w}^T \mathbf{x} = \mathbf{w}^T A \mathbf{s} = \mathbf{z}^T \mathbf{s} \quad (3.28)$$

Widzimy więc że y jest kombinacją liniową \mathbf{s} . Zgodnie z centralnym twierdzeniem granicznym suma dowolnych rozkładów prawdopodobieństwa dąży do rozkładu Gaussa, jeśli jest spełniony warunek Lindeberga, a w szczególności gdy $E[X] = 0$ i $\sum_{k=1}^n \text{var}[X] = 1$ to:

$$\sum_{k=1}^n X_k \rightarrow N(0, 1) \quad (3.29)$$

Widzimy więc że \mathbf{y} będące kombinacją liniową wektorów \mathbf{s} , będzie bardziej "gaussowskie", niż dowolna z niezależnych składowych osobno. Założyliśmy że \mathbf{s} nie ma rozkładu Gaussa, więc \mathbf{y} będzie równe \mathbf{s} , gdy będzie najmniej gaussowskie, co jest jednoznaczne z tylko jedną niezerową wartością w wektorze \mathbf{z} .

Autorzy [9] proponują miary niegaussowskości opartą na kurtozie, bądź ujemnej entropii. Negatywna entropia może być dość dobrym estymatorem stopnia w jakim rozkład nie jest gaussowski, jest nieujemna i równa zero tylko dla rozkładu Gaussa.

Oprócz zastosowania w analizie mowy, często podowaną aplikacją ICA jest ocenianie aktywności mózgu w poszczególnych obszarach na podstawie sygnałów z elektrod rozmieszczonych w różnych miejscach głowy. W kontekście tej pracy zależenie wektorów \mathbf{s} , mogących definiować bazę nowej przestrzeni prowadzi do redukcji wymiarowości danych.

Rozdział 4

Zabezpieczenie przed możliwością oszukania systemu

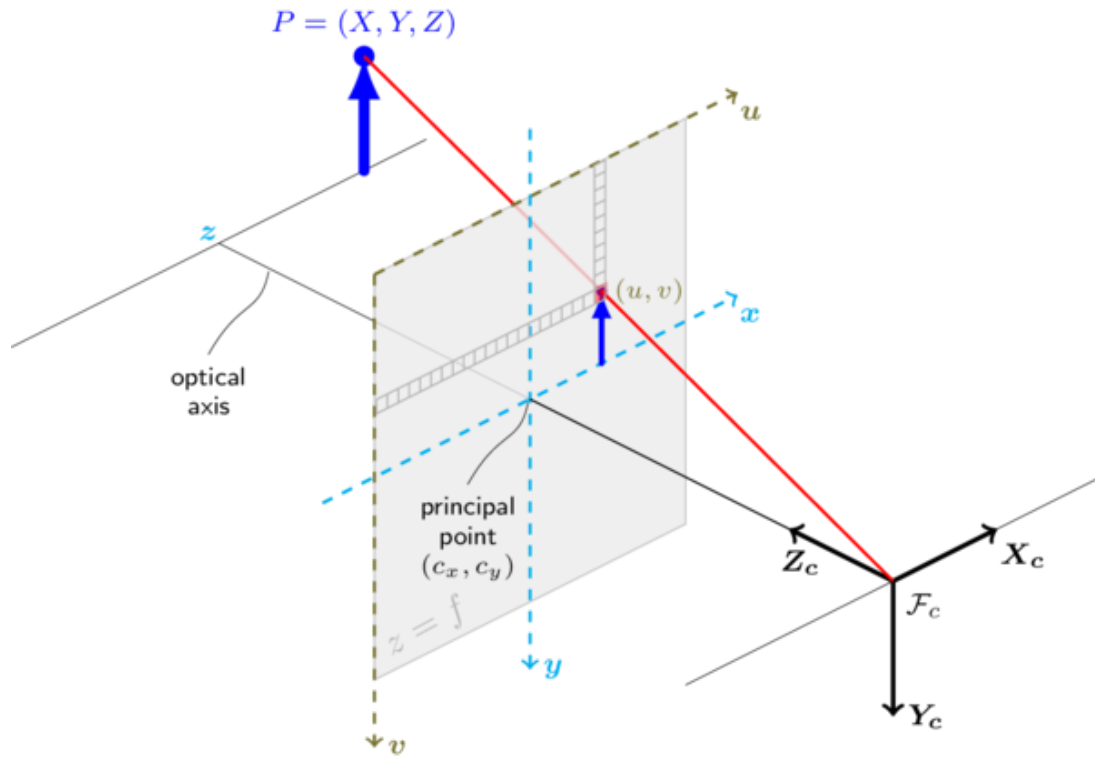
W trakcie realizacji projektu specjalnościowego podjęto próbę zaadresowania kwestii oszukiwania całego systemu poprzez przedstawianie systemowi zdjęć ludzi, którym wcześniej przyznano prawo dostępu. Jest to realne zagrożenie, z którym twórcy systemów biometrycznych muszą się mierzyć. Dobrym przykładem może być tutaj system FaceID w iPhoneX firmy Apple. Twórcy systemu zapewniali że nie da się oszukać poprzez wykonanie masek 3d imitujących twarz człowieka. 9 listopada 2017 roku eksperci z firmy Bkav oświadczyli że udało im się złamać zabezpieczenia poprzez wykonanie specjalnej maski, wykorzystując drukarkę 3D, silikonowe elementy, wydruki 2D, oraz materiał imitujący skórę [15]. Dodatkowo system FaceID okazał się nie być odporny na podobieństwo występujące między bliźniakami jednojajowymi. Bezpieczeństwo systemów biometrycznych jest więc bardzo ważną kwestią i podjęto próbę zbadania tych zagadnień.

4.1 Przyjęta metodyka

Zbadano możliwości jakie daje ekstrakcja głębi z wykorzystaniem dwóch kamer. W obecnym projekcie systemu zintegrowanie systemu ekstrakcji głębi byłoby trudne, ponieważ w projekcie wykorzystywana jest kamera Raspberry Pi, podłączona przez port CSI, a mikrokomputery Raspberry Pi posiadają tylko jeden port CSI. Do badań wykorzystano dwie kamery internetowe Creative VF0770. Do ekstrakcji głębi wykorzystano funkcje biblioteki OpenCV. Celem było uzyskanie mapy punktów w 3D, które następnie miały być aproksymowane płaszczyzną. Następnie należało wyznaczyć błąd aproksymacji, oraz próg, od którego mapa punktów miała być uznawana za płaską.

4.2 Podstawy teoretyczne

Przed przedstawieniem wyników omówiono przyjęte założenia teoretyczne. W niniejszej sekcji zostały omówione takie koncepcje jak pinhole camera model, model zniekształceń w bibliotece OpenCv, podstawowe założenia geometrii epipolarnej oraz wyznaczanie mapy



Rysunek 4.1: Scena przy przyjętym modelu kamery.

Źródło:

https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

głębokości na podstawie dysparycji.

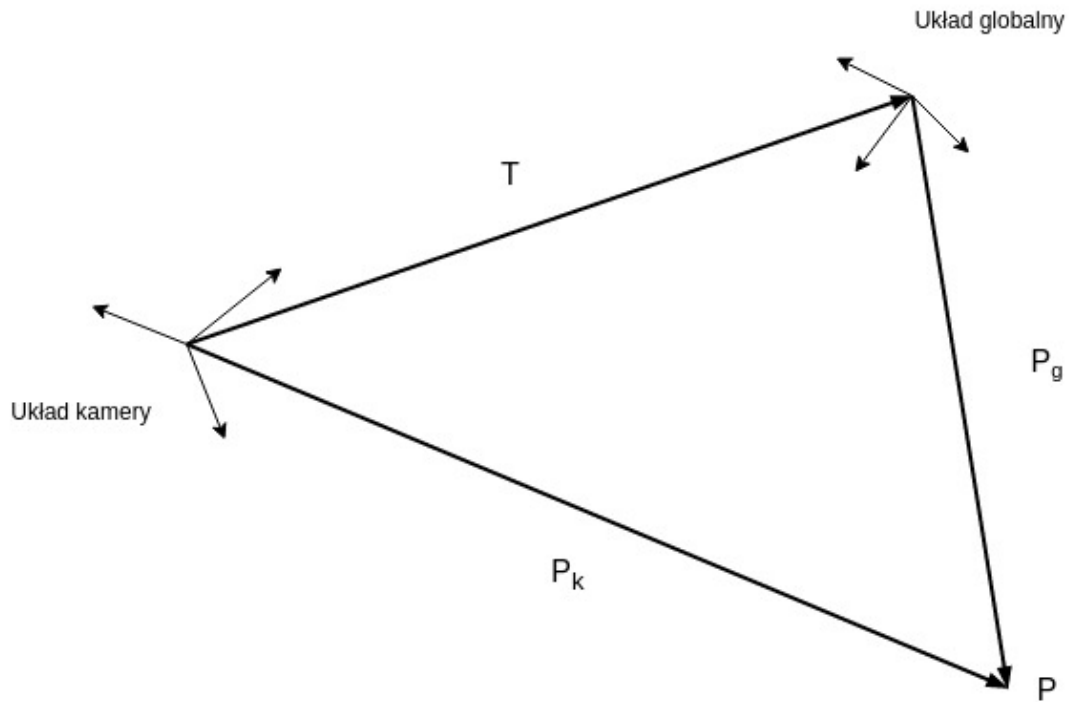
Model kamery

Przed przystąpieniem do ekstrakcji głębi należy ustalić model działania pojedynczej kamery oraz zrozumieć jakie zakłócenia obrazu występują.

Najczęściej rozpowszechnionym modelem jest tzw. pinhole camera model. W modelu tym kamera przekształca piksele z globalnego układu współrzędnych do układu współrzędnych na płaszczyźnie, na którą są rzutowane punkty. Przekształcenie to można wyrazić wzorem:

$$Z \begin{pmatrix} u \\ v \end{pmatrix} = K(R \ T) \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (4.1)$$

W powyższym równaniu X, Y oraz Z to współrzędne punktu w zewnętrznym układzie współrzędnych. u i v to współrzędne w układzie kamery. Macierz R i wektor T to macierz rotacji i wektor translacji opisujące orientację i położenie kamery względem zewnętrznego układu współrzędnych. Zaliczają się one do parametrów zewnętrznych, ponieważ nie zależą



Rysunek 4.2: Ilustracja zależności między układami współrzędnymi a punktem.

od samej kamery, ale od jej położenia i orientacji. Parametry te opisują transformację punktu z globalnego układu współrzędnych do układu współrzędnych kamery.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (R \quad T) \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Gdzie x, y, z to współrzędne punktu P w układzie kamery.

Warto zwrócić uwagę że wektor T zawiera położenie globalnego układu współrzędnych względem układu współrzędnych kamery, a nie odwrotnie. Fakt ten może wydawać nieco nieintuicyjny, ale można go łatwo zilustrować. Z 4.2 wynika jednoznacznie że $P_k = T + P_g$. Oczywiście należy uwzględnić wzajemną orientację układów, co zapewnia macierz R .

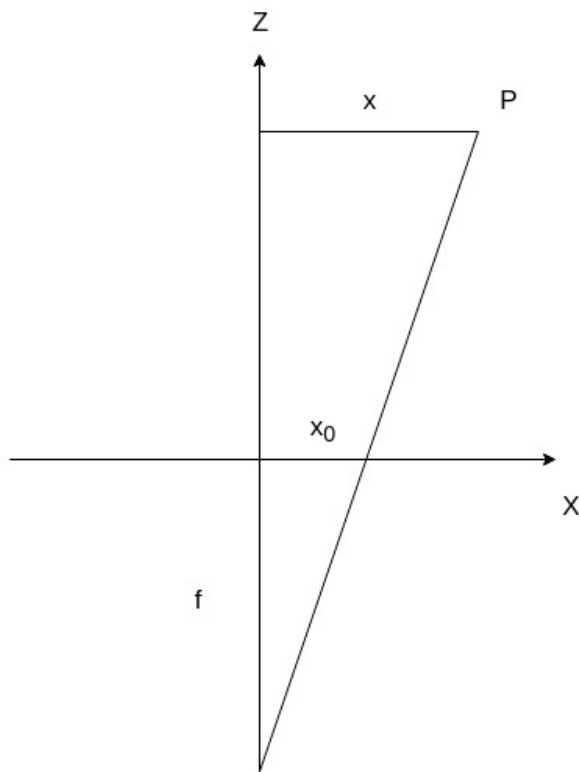
Macierz K jest zawiera parametry wewnętrzne kamery i jest definiowana jako:

$$K = \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (4.2)$$

Gdzie f to długość ogniskowej, c_x, c_y jest punktem w którym oś optyczna przecina się płaszczyzną obrazu (ang. principal point), zwykle znajdujący się w środku obrazu.

Korzystając z wcześniejszych równań można zapisać:

$$\begin{aligned} u &= f \frac{x}{z} + c_x \\ v &= f \frac{y}{z} + c_y \end{aligned}$$



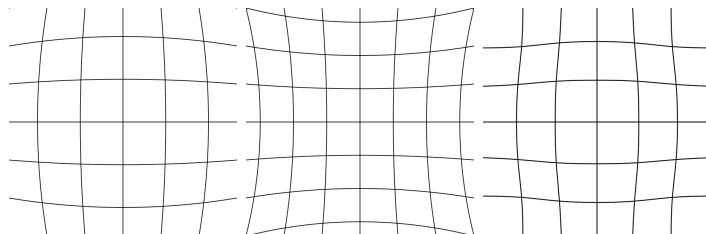
Rysunek 4.3: Rzut z góry na scenę.

Na rys 4.3 można zauważyć że równania te mają proste uzasadnienie geometryczne. Z twierdzenia Talesa otrzymujemy: $\frac{x_0}{f} = \frac{x}{z}$. Wartość u możemy uzyskać poprzez dodanie do x_0 z poprzedniego równania c_x . Dla y rozumowanie jest analogiczne.

Zakłócenia

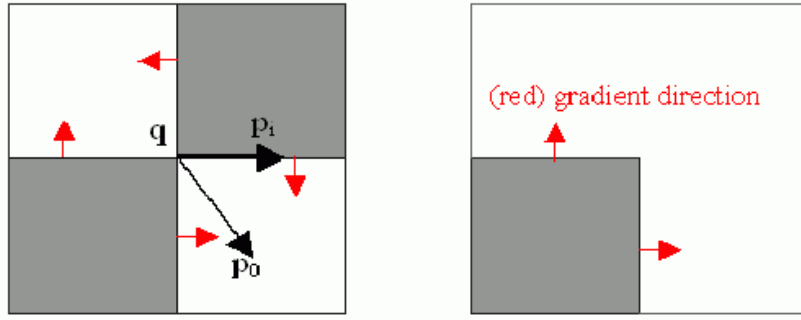
Zaprezentowany wcześniej model zakłada brak zakłóceń przy procesie przetwarzania obrazu. W kamerach występują różne rodzaje dystorsji. W dystorsji nie linie proste zostają zniekształcone. Wynika z fizycznej niedoskonałości układów optycznych. Im dalej od osi optycznej tym większe zniekształcenie.

W bibliotece OpenCV przewiduje się współczynniki radialne k_1, \dots, k_6 oraz styczne p_1 i p_2 . Współczynniki te są wykorzystywane w poprzednim modelu w następujący sposób



Rysunek 4.4: Różne rodzaje dystorsji (od lewej): beczkowa, poduszkowa oraz falista.

źródło: https://en.wikipedia.org/wiki/Distortion_%28optics%29



Rysunek 4.5: Gradient szachownicy.

źródło: https://docs.opencv.org/2.4/modules/imgproc/doc/feature_detection.html

(podawane za dokumentacją OpenCV [12]):

$$\begin{aligned}
 x' &= \frac{x}{z} \\
 y' &= \frac{y}{z} \\
 x'' &= x' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + 2p_1 x' y' + p_2 (r^2 + 2x'^2) \\
 y'' &= y' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_4 r^2 + k_5 r^4 + k_6 r^6} + p_1 (r^2 + 2y'^2) + 2p_2 x' y'
 \end{aligned}$$

4.2.1 Kalibracja kamery

Kalibracja kamery odbywa się poprzez fotografowanie szachownicy pod różnymi kątami. Następnie zdjęcia są konwertowane do skali szarości i wykrywano narożniki szachownicy. Pierwsze oszacowanie położenia narożników nie jest dokładne. Estymacja jest poprawiana z wykorzystaniem prostego faktu odnośnie gradientu:

Założmy że \mathbf{q} jest narożnikiem szachownicy, a \mathbf{p}_i jest punktem w bliskim sąsiedztwie. Przyjmijmy wartość błędu daną następującym równaniem:

$$E_i = \nabla f_{\mathbf{p}_i}^T (\mathbf{q} - \mathbf{p}_i) \quad (4.3)$$

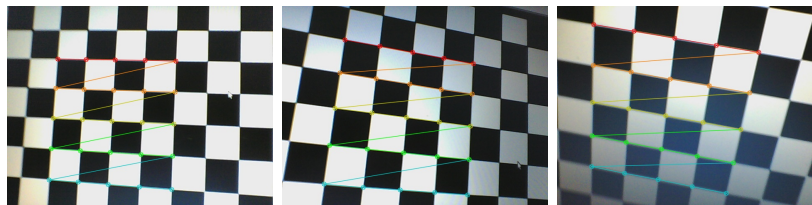
Gdzie $\nabla f_{\mathbf{p}_i}$ jest gradientem, w punkcie \mathbf{p}_i . Jest to równanie na iloczyn skalarny i będzie ono równe zero, gdy \mathbf{p}_i będzie prostopadłe do gradientu. Przyrównując powyższe równanie do zera, oraz mnożąc obie strony przez $\nabla f_{\mathbf{p}_i}$ otrzymujemy:

$$\nabla f_{\mathbf{p}_i} \nabla f_{\mathbf{p}_i}^T \mathbf{q} - \nabla f_{\mathbf{p}_i} \nabla f_{\mathbf{p}_i}^T \mathbf{p}_i = 0$$

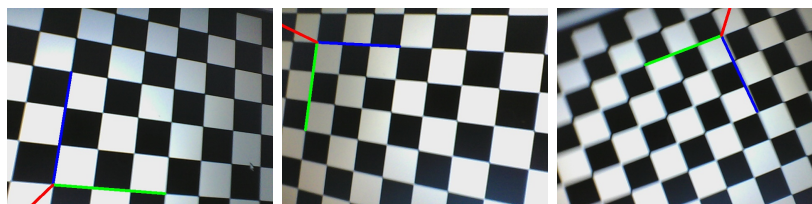
W konsekwencji otrzymujemy:

$$\mathbf{q} = (\nabla f_{\mathbf{p}_i} \nabla f_{\mathbf{p}_i}^T)^{-1} \nabla f_{\mathbf{p}_i} \nabla f_{\mathbf{p}_i}^T \mathbf{p}_i \quad (4.4)$$

Operacja ta jest powtarzana, aż punkt q uzyska stabilną wartość (jego współrzędne nie będą się zmieniać o większą wartość niż przyjęty próg).



Rysunek 4.6: Wykryte narożniki szachownicy.



Rysunek 4.7: Estymacja orientacji szachownicy.

Dzięki wyznaczonym narożnikom można wyznaczyć wewnętrzne i zewnętrzne kamery. Wartości te są następnie wykorzystane do estymacji pozycji 4.7, oraz do usunięcia zakłóceń z obrazu.

4.2.2 Geometria epipolarna

Zakładając wcześniej przyjęty model rzeczywistej kamery oraz stosując uproszczenie polegające na przeniesieniu płaszczyzny obrazu przed ognisko (w rzeczywistych kamerach płaszczyzna ta znajduje się za ogniskiem) możemy zauważyć pewne zależności geometryczne dla punktu widzianego z dwóch kamer.

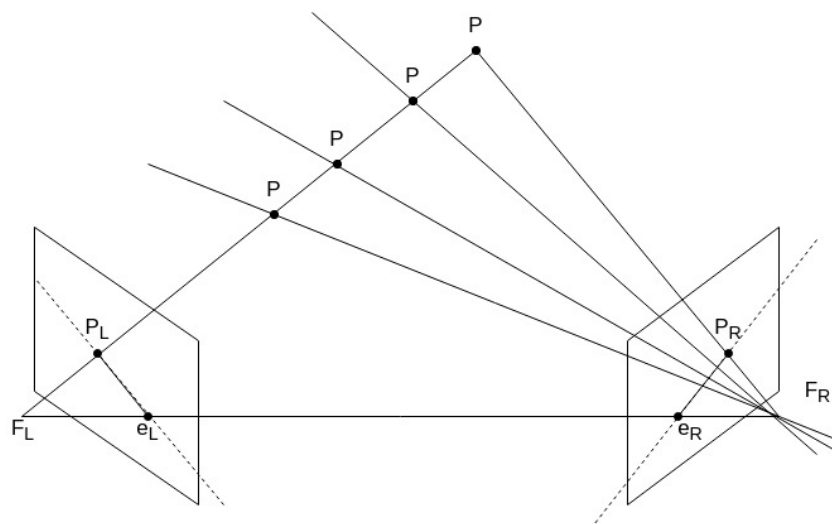
Pierwszym spostrzeżeniem jest fakt że ogniska obu kamer, punkt P oraz jego projekcje na płaszczyznę obrazu leżą na jednej płaszczyźnie. Linia poprowadzona pomiędzy ogniskami kamer jest nazywana linią główną (ang. baseline). Punkty przecięcia linii głównej z płaszczyzną główną (e) nazywane są epipolami (ang. epipoles). Jeśli obie kamery się widzą to epipole są widoczne jako środek drugiej kamery na obrazie. Płaszczyzna przechodząca przez dowolny punkt P widziany w obu kamerach, oraz przez linię główną jest nazywana płaszczyzną epipolarną. Linia epipolarną nazywamy linią będącą na przecięciu płaszczyzny obrazu z płaszczyzną epipolarną. Wszystkie linie epipolarne przecinają się w epipolach.

Warto zwrócić uwagę na to że linia FP jest widziana na drugim obrazie jako odpowiadająca linia epipolarna. Oznacza to więc, że szukając odpowiednika piksela z jednego obrazu na drugim, należy szukać wzdłuż epilinii, a nie na całym obrazie. Jest to znaczące ułatwienie, które będzie miało duże znaczenie przy obliczaniu dysparycji.

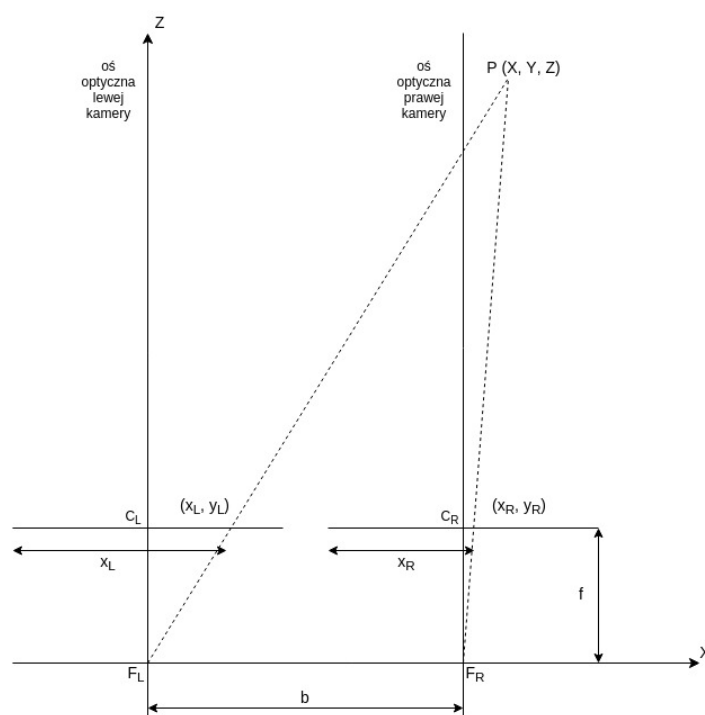
4.2.3 Szacowanie głębokości

Kolejnym krokiem jest obliczenie mapy dysparycji oraz na jej podstawie wyznaczenie głębokości.

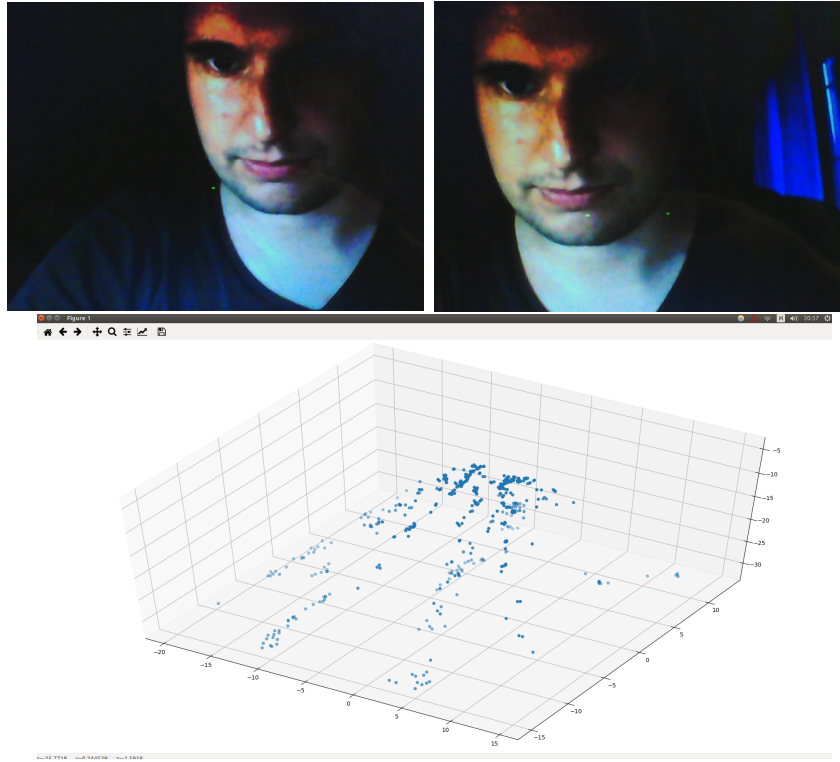
Przy ekstrakcji głębokości jest wykorzystywane podobieństwo trójkątów PF_LF_R oraz PC_LC_R z rys. 4.9:



Rysunek 4.8: Podstawowa koncepcja geometrii epipolarnej.



Rysunek 4.9: Ekstrakcja głębi na podstawie dysparycji obrazów z dwóch kamer.



Rysunek 4.10: Obraz płaskiego zdjęcia sfotografowany lewą i prawą kamerą oraz wynik ekstrakcji głębokości.

$$\frac{b}{Z} = \frac{b - x_L + x_R}{Z - f} \quad (4.5)$$

Po przekształceniach otrzymujemy:

$$Z = \frac{bf}{x_L - x_R} = \frac{bf}{d} \quad (4.6)$$

Gdzie d jest dysparycją - różnicą wartości odpowiadających sobie pikseli na dwóch obrazach, szukanych wzdłuż epilinii. Znając ogniskową, odległość między kamerami, oraz obliczając różnicę między analogicznymi pikselami z dwóch obrazów możemy szacować głębie.

Warto zwrócić uwagę że wzór 4.5 nie traci na ogólności, gdy punkt P znajdzie się pomiędzy osiami optycznymi kamer, ponieważ do obliczenia podstawy trójkąta $PC_L C_R$ zostanie wykorzystana odległość między punktem w którym wiązka światła pada na prawą matrycę a osią optyczną prawej kamery. Odległość ta może zostać w takim wypadku wyrażona jako $\frac{1}{2}l - x_R$, gdzie l to długość ekranu światłoczułego. Po podstawieniu podobnej wartości za odległość od osi optycznej lewej kamery do punktu przecięcia wiązki światła z lewą matrycą $\frac{1}{2}l$ się uprości i zostaniemy ze wzorem 4.5.

4.3 Wyniki

Dokonano kalibracji kamery, wyznaczono współczynniki wewnętrzne i zewnętrzne kamery. Jednakże nie udało się uzyskać zadowalających rezultatów. Po przedstawieniu systemowi

zdjęcia osoby wyświetlonego na laptopie uzyskano mapę głębi przedstawioną na rys. 4.10. Jak można z łatwością zauważyć na mapie głębokości otrzymano punkty o różnej wartości współrzędnej z.

Trudno jest określić jedną przyczynę niepowodzenia. Na pewno ważną rolę odgrywa kalibracja kamer. Wielokrotnie podejmowano próbę poprawiania uzyskiwanych wyników poprzez ponowną kalibrację, ale nie wpłynęło to w znaczący sposób na rezultaty. W związku z ograniczoną ilością czasu projekt nie mógł zostać ukończony, pomimo wielu godzin poświęconych na analizę problemu.

Rozdział 5

Zakończenie

5.1 Podsumowanie projektu

W trakcie realizacji projektu przeanalizowano dostępne metody detekcji twarzy i ekstrakcji cech. Następnie wybrano algorytmy adekwatne dla realizowanego projektu i wykorzystano je. Bardzo pomocne były biblioteki, które pozwoliły skupić się na tworzeniu systemu, a nie implementacji algorytmów. Oszczędziło to czas potrzebny na implementację rozwiązań. W projekcie przyjęto rozwiązanie polegające na detekcji twarzy i wykorzystaniu algorytmów redukcji wymiarowości do przygotowania wektora uczącego dla sieci o wielkości 400 elementów. Jest to dość popularne rozwiązanie tego problemu i daje dość dobre rezultaty. Nie przeprowadzono testów modułowych, system był testowany jako całość. Powodem takiego stanu rzeczy jest trudność w testowaniu omawianego komponentu, ze względu na dość abstrakcyjny charakter, w porównaniu do reszty abstrakcji. Testowanie rezultatów działania całego systemu jest kosztowne, ponieważ dowolna zmiana w systemie ekstrakcji cech, powoduje konieczność zaczynania procesu uczenia sieci neuronowej od nowa, co jest kosztowne i czasochłonne. W dalszej części tego rozdziału przedstawiono koncepcję testów modułowych dla komponentu realizującego ekstrakcję cech.

Realizowany projekt umożliwił także opracowanie i wykorzystanie wydajnych sposobów na pracę w grupie. Komponentowe podejście do systemu było bardzo dobrym rozwiązaniem, umożliwiającym zrównoleglenie pracy. Dodatkowo ustalenie interfejsów między komponentami pozwala na przeprowadzanie daleko idących zmian w obrębie komponentu, bez konieczności zmiany innych części systemu. Przeprowadzenie refaktoryzacji kodu po spełnieniu podstawowych założeń systemu było bardzo pomocne i ułatwiło rozwój aplikacji.

5.2 Możliwe usprawnienia

System w obecnym stanie spełnia podstawowe założenia jakie zostały przed nim postawione, t.j. rozpoznawanie twarzy niewielkiej grupy osób oraz ich identyfikacja. Jednakże jest wiele aspektów które można poprawić i rozwinąć. W niniejszej sekcji przeprowadzono dyskusję najważniejszych problemów, które należy zaadresować przy pracy nad projektem.

Testy modułowe

Jednym z pomysłów na rozwój systemu jest porównanie efektów działania algorytmów redukcji wymiarowości. Metodyka takich testów miałaby polegać na utworzeniu bazy zdjęć różnych osób i porównanie odległości współrzędnych przyporządkowanych danym zdjęciom z bazy przez różne algorytmy. Możliwe do analizowania miary to: średnia odległość, największa i najmniejsza odległość, mediana oraz wariancja. Przy takim porównaniu należy oczywiście wybrać odpowiedni rodzaj normy. Takie podejście umożliwiłoby uniezależnienie wyników omawianego komponentu od sieci neuronowej, co znacznie ułatwiłoby badania właściwości proponowanych rozwiązań. Warto zwrócić uwagę że takie rozwiązanie pozwala na testowanie dowolnego podejścia do problemu. Ustalenie wspólnego zbioru zdjęć dla porównywanych algorytmów, normy dla przestrzeni cech, oraz miar i progów dla każdej z miar umożliwiłoby automatyzację testów.

Zbadanie wpływu czynników wizualnych na wyniki systemu

Mając dostęp do usystematyzowanych testów można dokładnie zbadać wyniki jakie osiąga system oraz jego odporność na czynniki wizualne. Do takich czynników zalicza się: różne ustawienie twarzy, oświetlenie kierunkowe twarzy, pory dnia, zarost czy okulary. Proponuje się przyjęcie zbioru zdjęć jednej osoby w różnych warunkach oświetleniowych i porównanie uzyskanych wektorów.

Geometryczne cechy

W rozdziale 2.2 przedstawiono koncepcję wykorzystania innego rodzaju cech. Polega na przygotowania wektora cech z geometryczną interpretacją. Każdy element takiego wektora mógłby oznaczać odległość między charakterystycznymi punktami twarzy. Aby uniezależnić uzyskiwane wyniki od odległości wszystkie odległości muszą być względne. W przypadku tego typu wektora należy przetestować jak sieć neuronowa poradzi sobie z tego typu cechami, oraz zbadać wpływ pozycji twarzy na osiągane wyniki. Algorytmy estymujące położenie punktów charakterystycznych z reguły potrafią sobie poradzić z oświetleniem, więc nie byłby to czynnik mający wielkie znaczenie w tym przypadku, można go jednakże dodać do testów, dla podwójnego sprawdzenia.

Ekstrakcja głębi

Ekstrakcja głębi na podstawie obrazów z dwóch kamer w obecnym stanie projektu nie daje zadowalających rezultatów. Fakt ten prawdopodobnie jest spowodowany złą kalibracją kamery, bądź niewłaściwym położeniem, bądź określeniem położenia kamer. Jest to część projektu, która musi zostać dopracowana, następnie może zostać zintegrowana do systemu. Poza zwiększeniem bezpieczeństwa systemu można wykorzystać mapę głębi do ekstrakcji cech. Ponieważ taka mapa ma zbyt wiele punktów konieczne byłoby wykorzystanie algorytmów redukcji wymiarowości. Bardzo ciekawe mogłoby być porównanie wyników uzyskiwanych przez system na podstawie obrazów 2D i 3D. Informacje uzyskane dzięki estymacji głębi otwierają bardzo wiele możliwości rozwoju systemu.

Bibliografia

- [1] Christopher M. Bishop. *Pattern recognition and machine learning*. Springer Science + Business Media, New York 2009
- [2] Krzysztof Ślot. *Rozpoznawanie biometryczne. Nowe metody ilościowej reprezentacji obiektów* Wydawnictwo Komunikacji i Łączności, Warszawa 2000
- [3] Stanisław Osowski. *Sieci Neuronowe w ujęciu algorytmicznym* Wydawnictwo Naukowo-Techniczne, Warszawa 1996
- [4] Yoav Freund, Robert E. Schapire. *A Short Introduction to Boosting* Journal of Japanese Society for Artificial Intelligence, 1999
- [5] Paul Viola, Michael Jones. *Rapid Object Detection using a Boosted Cascade of Simple Features*. Cambridge 2001
- [6] Vahid Kazemi, Josephine Sullivan. *One Millisecond Face Alignment with an Ensemble of Regression Trees* KTH, Royal Institute of Technology Computer Vision and Active Perception Lab, 2014
- [7] Daniel D. Lee and Seung, H. Sebastian. *Advances in Neural Information Processing Systems 13: Algorithms for Non-negative Matrix Factorization* str. 556-562 MIT Press, 2001
- [8] C. Boutsidis, E. Gallopoulos. *SVD based initialization: A head start for nonnegative matrix factorization* University of Patras 2007
- [9] Aapo Hyvärinen, Erkki Oja *Independent Component Analysis: Algorithms and Applications* Helsinki University of Technology, 2000
- [10] Hartley, R. I., Zisserman, A. *Multiple View Geometry in Computer Vision* Cambridge University Press, 2004
- [11] Itseez. *Open Source Computer Vision Library*
<https://github.com/itseez/opencv>
- [12] Dokumentacja biblioteki OpenCv
<https://docs.opencv.org/2.4/index.html>
- [13] Zbiór tutoriali z OpenCV
https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_tutorials.html

- [14] Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E. *Scikit-learn: Machine Learning in Python* Journal of Machine Learning Research, volume 12, strony 2825-2830, 2011
- [15] News o pokonaniu zabezpieczeń systemu FaceID
http://www.bkav.com/d/topnews//view_content/content/103968/faceid-beatenbymasknotaneffectivesecuritymeasure

Spis treści

1	Wstęp	1
1.1	Omówienie zagadnienia	1
1.2	Omawiany komponent jako część większego systemu	1
1.2.1	Wykorzystane wzorce projektowe	2
1.2.2	Wykorzystane zasady i dobre praktyki programowania	2
1.2.3	Wykorzystane biblioteki, narzędzia i zasoby	2
2	Implementacja algorytmów – detekcja twarzy	4
2.1	Kaskady Haara	4
2.2	Inne sposoby rozwiązania problemu	6
3	Implementacja algorytmów – ekstrakcja cech	8
3.1	Principle Components Analysis	8
3.1.1	Wstęp matematyczny	8
3.1.2	Realizacja algorytmu	11
3.1.3	Interpretacja	12
3.1.4	PCA jako samodzielny system rozpoznawania twarzy	13
3.2	Inne podejścia do problemu redukcji wymiarowości	14
3.2.1	Nonnegative Matrix Factorization	14
3.2.2	Independent Components Analysis	15
4	Zabezpieczenie przed możliwością oszukania systemu	17
4.1	Przyjęta metodyka	17
4.2	Podstawy teoretyczne	17
4.2.1	Kalibracja kamery	21
4.2.2	Geometria epipolarna	22
4.2.3	Szacowanie głębokości	22
4.3	Wyniki	24
5	Zakończenie	26
5.1	Podsumowanie projektu	26
5.2	Możliwe usprawnienia	26

Spis rysunków

2.1	Różne rodzaje cech. Od sumy wartości pikseli po czarnym obszarem jest odejmowana wartość sumy pikseli pod białym polem.	5
2.2	Dwie cechy generujące najmniejszy błąd z wykorzystaniem AdaBoost. . . .	5
2.3	Wykrywanie punktów charakterystycznych twarzy.	6
3.1	Porównanie rekonstrukcji twarzy na podstawie różnej liczby wykorzystanych składowych.	13
3.2	Porównanie eigenfaces.	13
4.1	Scena przy przyjętym modelu kamery.	18
4.2	Ilustracja zależności między układami współrzędnymi a punktem.	19
4.3	Rzut z góry na scenę.	20
4.4	Różne rodzaje dystorcji (od lewej): beczkowa, poduszkowa oraz falista. . .	20
4.5	Gradient szachownicy.	21
4.6	Wykryte narożniki szachownicy.	22
4.7	Estymacja orientacji szachownicy.	22
4.8	Podstawowa koncepcja geometrii epipolarnej.	23
4.9	Ekstrakcja głębi na podstawie dysparycji obrazów z dwóch kamer.	23
4.10	Obraz płaskiego zdjęcia sfotografowany lewą i prawą kamerą oraz wynik ekstrakcji głębokości.	24