Politechnika Warszawska

WYDZIAŁ ELEKTRYCZNY

Instytut Elektrotechniki Teoretycznej
i Systemów Informacyjno-Pomiarowych
Zakład Elektrotechniki Teoretycznej
i Informatyki Stosowanej

Praca dyplomowa inżynierska

na kierunku Informatyka
w specjalności Inżynieria oprogramowania

Rozpoznawanie gestów na przykładzie gry papier, kamień, nożyce

Adrian Szewczyk
nr albumu 279074

promotor
mgr inż. Marek Wdowiak

WARSZAWA 2018

Rozpoznawanie gestów na przykładzie gry papier, kamień, nożyce

Streszczenie

W pracy zostało przedstawione podejście do problemu rozpoznawania gestów przez wszystkie jego etapy. Na początku zostało pokazane zabranie i przetwarzanie danych w celu uzyskania jak najlepszego zbioru uczącego jak i testowego. Potem zostały opisane etapy filtracji, segmentacji dłoni. Po uzyskaniu danych treningowych został stworzony klasyfikatory. Pierwszym podejściem było stworzenie klasyfikatora KNN, a następnie stworzenie klasyfikatora bazującego na sieciach neuronowych. Końcowym etapy pracy to porównanie wyników uzyskanych przez wymienione klasyfikatory na przykładzie gry papier, kamień, nożyce.

Słowa kluczowe: rozpoznawanie gestów, klasyfikacja, TensorFlow, sieci neuronowe

ENG Rozpoznawanie gestów na przykładzie gry papier, kamień, nożyce

Abstract

Keywords: ENG rozpoznawanie gestów, klasyfikacja, python, sieci neuronowe

WARSZAWA, 1 lutego 2018

POLITECHNIKA WARSZAWSKA
WYDZIAŁ ELEKTRYCZNY

OŚWIADCZENIE

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa inżynierska pt. Rozpoznawanie gestów na przykładzie gry papier, kamień, nożyce:

- została napisana przeze mnie samodzielnie,
- nie narusza niczyich praw autorskich,
- nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam, że przedłożona do obrony praca dyplomowa nie była wcześniej podstawą postępowania związanego z uzyskaniem dyplomu lub tytułu zawodowego w uczelni wyższej. Jestem świadom, że praca zawiera również rezultaty stanowiące własności intelektualne Politechniki Warszawskiej, które nie mogą być udostępniane innym osobom i instytucjom bez zgody Władz Wydziału Elektrycznego.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Adrian Szewczyk.....

Spis treści

Rozdział 1

Wstęp

1.1 Wprowadzenie

Gesty są jednym z najważniejszych elementów komunikacja pomiędzy ludźmi. Niewerbowana mowa może przekazać zdecydowanie więcej informacji niż tylko słowa wypowiedziane przez nadawcę. Gesty mogę, również zastąpić słowa i być wykorzystywane jako zamiennik na słyszalne słowa. Przykładem tego jest język migowy, który pozwala na wzajemną komunikację ludzi głuchych . Oprócz samej mowy gesty, mają też inne zastosowania. Przykładem takim może być gra papier, kamień, nożyce. W której każdy z graczy pokazuje jeden z trzech dostępnym gestów. I w zależności od kombinacji gestów pokazanych przez graczy określany jest zwycięzca.

Klasyfikacja gestów jest powszechnym problem przetwarzania obrazów. Jest to temat obszerny i coraz lepiej zbadany problem nauki. Klasyfikacja gestów ma szerokie zastosowanie. Jedną zalet rozpoznawania gestów jest możliwość sterowania naszym smartphonem za pomocą ruchu rąk. Po wykonaniu gestu a następnie po rozpoznaniu go przez oprogramowanie telefonu wykonywana jest określona akcja. Innym wykorzystaniem klasyfikacji gestów jest sterowanie gier, a najlepszym jego przykładem jest konsola Nintendo Wii.

1.2 Cel pracy

W swojej pracy dypmowej będę chciał przedstawić w jaki sposób można modejść do klasyfikacji gestów. Zaimplementować konkretne, działające rozwiązanie i wykorzystać go w prakce czym będzie gra papier,kamień, nożyce bez konieczności ręcznego liczenia punktów i byciem arbitrem wszystkim tym zajmie się specjalny system, który będę chciał zaprezentować w tej pracy dyplomowej.

Rozdział 2

Teoria

2.1 Przetwarzanie obrazów

2.1.1 Cyfrowe przetwarzanie obrazów binarnych

dziedzina przetwarzania obrazów cyfrowych zajmująca się algorytmami obróbki obrazów binarnych. Obrazy takie składają się z pikseli, które mogą przyjmować jedynie dwie wartości. Piksele takie zaś można oznaczać róznymi symboli, np. 0/1, false/true xzy też (0,0,0)/(255,255,255). Podczas interpretacji obrazu jedne z pikseli można traktować jako tło, inne zaś jako część obiektu. Dzięki czemu jesteśmy wstanie rozróżnić interesującą nas część obrazu od pozostałej części.

2.1.2 OpenCV

OpenCV (Open Source Computer Vision Library) jest to otwartoźródłowa biblioteka wykorzystywana przy rozpoznawaniu obrazów czy też uczeniu maszynowym. Została napisana w języku C przez programistów z firmy Intel w 1999 roku.

Późniejszym czasie kolejne części biblioteki były także pisane w jęzuku C++. OpenCV nie tylko jest wykorzystywany w językach takich jak C/C++ ale również mamy możliwość wykorzystywania tej biblioteki w językach Python, Java czy też Matlab. Zostały różnież udostępnione nakładki do języków takich jak C#, Perl, Ruby czy Haskell aby móc również wykorzystywać zalety tej właśnie biblioteki.

OpenCV w szeroko rozumianym pojęciu rozpoznawaniu przetwarzaniu obrazów umożliwia zaawansowaną funkcjonalność w tematach: -przetwarzania obrazów -klasyfikowaniu wzorców -dokładnych pomiarów obrazów

Jako główne zalety OpenCv możemy wyróżnić, że jest jest to darmowa, otwartoźródłowa biblioteka. zabiera bogaty zakres funkcjonalności. Kod biblioteki został napisany w sposób zoptymalizowanym, tak aby operacje wymogają dyżej mocy obliczeniowej czy działające w czasie rzeczywistym mogły wykonywać się możliwie jak najszybciej.

OpenCV znalazło zastoswanie w wielu dziedzinach naszego codziennego życia: -medycyna-robotyka -samochody autonomiczne -systemy anywłamaniowe -systemy zabeczpiecające -rozpoznawanie gestów -segmentacja obiektów -wykrywanie ruchu -rozszerzona rzeczywistość -rozpoznawanie obiektów

2.1.3 Operacje morfologiczne

Operacje morflogiczne to podstawowe operacje przetwarzania obrazów. Pozwalają na złożone operacje związane z analizą kształyu poszczególnych elementów obrazu oraz położenia względem siebie. W wyniku operacji struktura obiektu na obrazie zostaje zmieniona, w celu osiągnięcia określonych rezultatów. Operacje morfologiczne najczęściej stosuje się dla obrazów binarnych, dla których ś podstawowymi oraz jednymi z najważniejszych operacji. Dzięki nim jesteśmy wstanie wyszczególnić interesujące nas części czy też przefiltrować z części nas nieinteresujące. Postawowe operacje morfologiczne to: erozja, dylatacja, otwarcie oraz zamknięcie. Wymienione operacje można ze sobą łączyć, tworząc zaawansowane systemy analizy. Operacje morfologiczne modyfikują wartość pikseli biorąc pod uwagę wartości pikseli ich otaczających. Liczbę punktów otoczenia określa tzw. element strukturalny, który definije wartości i ich rozmieszenie w otoczeniu. Szablon strukturalny(element strukturalny) posiada jeden wyróżniony punkt, nazywany punktem centralnym. Najczęściej stosowanym elementem strukturalnym jest kwadrat o boku o nieparzystej liczbie pikseli, które wszystkie przyjmują wartość równą 1. Etapy przekształceń morfologicznych: 1. Szablon strukturalny jest przesuwany po całym obiekcie, tak aby punkt centralny szablonu był analizowanym pikselem 2. Następuje porównanie otoczenia analizowanego piksela z elementem strukturalnym 3. W zależności od stosowanej operacji morfologicznej wartośc analizowane piksela zmienia się lub też pozostaje bez zmian http://analizaobrazu.x25.pl/articles/19 Najczęściej stosowanymi elementemi strukturalnymi są kwadrat o boku o nieparzystej liczbie pikseli, które wszystkie przyjmują wartość równą 1 oraz element strukturalny, który aproksymuje swoim kształcem koło. Erozja: Dylatacja jest operacją zwężania i zmniejszania poprzez usunięcie pikseli granicznych. Usuwa wszystkie mniejsze obiekty, które możemy zainterpretować jako szumy.

Etapy operacji erozji: 1. Element strukturalny przesuwany jest interacynie po całym obiekcie, tak aby punkt centralny szablonu był analizowanym

pikselem 2. Porównuje się otoczenie analizowanego piksela z elementem strukturalnym 3. Jeżeli przynamniej jeden piksel z otoczenia objętego przez szablon strukturalny ma wartośc równą 0, to punkt centralny przyjmuje wartość 0. Jeżeli zaś taki przypadek nie wystąpił piksel centralny zachowuje swoją poprzednią wartość

Powyższe kroki możemy zapisać formalnie:

Poniższy rysunek przedstawia przykładowy wynik operacji erozji na obrazie binarnym:

Dylatacja: Dylatacja jest operacją rozszerzania i zwiększania. Pozwala na wypełnienie dziur binarnych w obiektach. Jeżeli dwa obiekty są położene blisko siebie, może nastąpić złączenie.

Etapy operacji erozji: 1. Element strukturalny przesuwany jest interacynie po całym obiekcie, tak aby punkt centralny szablonu był analizowanym pikselem 2. Porównuje się otoczenie analizowanego piksela z elementem strukturalnym 3. Jeżeli przynamniej jeden piksel z otoczenia objętego przez szablon strukturalny ma wartośc równą 1, to punkt centralny przyjmuje również przyjmuje wartośc 1. Jeżeli zaś wszystkie piksele z otoczenia piksela centralnego określonego przez element strukturalny mają wartość 0 to wówczas piksel centralny przyjmuje wartość 0.

Powyższe kroki możemy zapisać formalnie:

Poniższy rysunek przedstawia przykładowy wynik operacji dylatacji na obrazie binarnym:

Otwarcie: Operacja otwarcia morfologicznego jest definiowana przez połączenie metod erozji oraz dylatacji. Metoda ta wygładza obiekt, usuwa niechciane szumy, a w dodatku nie modyfikuje znacząco wielkości obiektów tak jak to jest w przypadku zastosowania operacji erozji czy dylatacji.

Etapy operacji otwarcia: 1. wykonywana jest operacja erozji na zadanym obrazie 2. Wykonowana jest operacja dylatacji na obraz, który jest wynikiem erozji z punktu 1.

Powyższe kroki możemy zapisać formalnie:

Poniższy rysunek przedstawia przykładowy wynik operacji dylatacji na obrazie binarnym:

Zamknięcie: Operacja zamknięcia łączy w sobie połączenie dwóch metod dylatacji oraz erozji, analogicznie jak było w przypadku operacji zamknięcia, jednakże w odwrotnej kolejności. W rezeltatacie uzykujemy połączenie obiektów o zbliżonych odległościach, wypełnienie dziur w obiektach, jednakże końcowy kształt obiektu zostaje w dużym stopniu zmieniony

Etapy operacji zamknięcia: 1. wykonywana jest operacja dylatacji na zadanym obrazie 2. Następnie wykonywana jest operacja erozja na obraz, który jest wynikiem dylatacji z punktu 1.

Powyższe kroki możemy zapisać formalnie:

Poniższy rysunek przedstawia przykładowy wynik operacji dylatacji na obrazie binarnym:

2.1.4 Filtry ruchu

Jest to powszechna technika generowanie maski binarnej zawierającej poruszające sie obiekty. Jak sama nazwa mówi BS background subtraction wykorzystuje różnicę z kolejnych klatek z kamery. Odejmowanie dwóch następujących po sobie klatego pozwala nam prześlidzieć które z cześci obrazów zostały zmienianione. Różnica dwóch pikseli da nam wartość niezerową tylko wtedy jeżeli a analizwanych klatkach wartość różnic się.

Rysunek działania:

OpenCv w swojej bibliotece posiada wiele gotowych rozwiązać to wykrywania obiektów w ruchu. Jako najczęsciej wykorzystywanych możemy wyróżnic następujące: -BackgroundSubtractorMOG -BackgroundSubtractorMOG2 -BackgroundSubtractorGMG

Przykładowy kod OpenCV:

Obrazki prezentujące BS:

Aby efekt był jeszcze lepszy, należy pozbyć się szumów, które mogą wystąpić podczas znalizy klatek. Są one spowodowane różnymi odbiciami, cieniami czy innymi niezauważlnymi ruchami. Wówczas należy wykorzystać operacje morfoloczne, które zostały opisane w rozdziale 2.22

2.1.5 Detekcja kolorów

2.2 Sztuczna inteligencja

2.2.1 Neuron

2.2.2 Sieci neuronowe

2.3 Uczenie maszynowe

2.3.1 Klasyfikatory

Klasyfikikację czy wyznaczanie klasy decyzyjnej do której należy nowy, nieznany dotąd obiekt. Motody klasyfikacje możemy podzielić na dwie kategorie. Pierwsza z nich jest klasykacją nadzorowaną, druga zaś to klasykacja nienadzorowana.

Klasykacja nadzorowana w momencie uczenia klasyfkator tzn. podawania zbioru danych treningowych muszą w sobie zawierać etykietę(atrybut

decyzyjny), która oznacza do jakiej klasy należy kokretny przypadek. Zaś w przypadku klasyfikacji niedadzorowanej nie posiadamy atrybutu decyzyjnego.

2.3.2 Klasyfikator KNN

Klasyfikator KNN jest przykładem klasyfikacji nadzorowanej. Zbiór treningowy zawiera w sebie cechy dla klasyfikatora(atrybutu) oraz atrubyt decyzyjny, który określa przynależność do klasy. Elementami wejścia do klasyfkatora jest zbiór uczący. Wyjściem jest klasyfikator, które możemy wykorzystujemy do określania atrubytów decyzyjnych nieznanych wcześniej obiektów.

Po utworzeniu klasykatora bazującego na zbierze uczącym szacowanie atrybutu decyzjnygo odbywa się w następujacych krokach: 1. Ustalamy wartość k 2. Znajdujemy k obiektów treningowych najbliższych naszemu obiektowi 3. Analizowany obiekt nalezy do klasy najliczniejszew w znalezionym w zbirze z podpunktu 2

Ocenianie odległości pomiędzy dwoma obiektami polega na umieszeniu obiektow w przestrzeni d-wymierowe, gdzie d opisuje ilośc atrybutow opisujących obiektu. Metryka miary może odbywać się w różny sposób. Najczesciej jest to miara euklidesowa, jednakze mozemy wokorzystac inne miary tj. Manhattana czy Minkowskiego.

Po znaleziu k najbliższych sąsiadów artruby decyzyjny przyjmuje wartość od najbardziej licznej klasy w wyznaczonym zbiorze.

Rysunek:

2.3.3 Dobór cech

2.3.4 Ocena cech

2.3.5 Deep learning

2.3.6 Python

Python: Jest to interpretowany, obiektowy język wysokiego poziomu. Używany jest w szerokiej gamie aplikacjach. Jest językiem ogólnego przeznaczenia co zoznacza, żże może zostać wykorzystany praktycznie do wszystkiego. Jest rozwijany jako projekt otwartoźródłowy. Pojawił się w roku 1991, autorem języka jest holenderski programista Guido van Rossum

Zalety używania Pythona: -prosta,czytelna, klarowna składnia, zmniejszająca ilość potrzebnych linijek kodu w programach -dynamicznie zarządza

pamięcią oraz typami danych -nie wymusza stylu programowania -jest wieloplatformowy -posiada bogaty zbiór różnegoradzaju bibliotek -łatwy do nauczenia, nawet dla osób zaczynajacych programować -szybkość działania w stosunku do innych języków kryptowych

Jeśli mielibyśmy dyskutować o bwadach językach Python to ciężko jednoznacznie wskazać takie. Na pewno część z programistów może zalety danamicznego zarządzania typami unać jako wadę, poniewaqż w niektórych przypadkach może spowodować błedy trudniejsze do znalezienia. Działania Pythona jest również wolniejsze do jęzków takich jak C czy C++. Za jedną z wad na pewno można uznać sposób programonia obiektowego jakie odbywa się w Pythonie. W szczególności nie ma enkapsulacji, istnieją metedy które symulują takie działa, jednakże w stosunku do innych języków obiektowych czytolność kodu zdecydowanie jest zmniejszona

Język mimo swoich lat ciągle zyskuje na popularności. Ostatnio wszet do pierwszej trójki najbardziej popularnych języków programowania według TIOBE(stan na grudzień 2018), wyprzedzając między innymi języki C+, a ulagając jedynie języką Java oraz C. Jeżyk Python na pewno wzrost poopularności łaczy się ze wzrostem popularności maching learningu i w tym deep learningu gdzie język Python jest jesdnym z najslepszych jak nie najlepszymn jeżykiem w tych dziedzinach. , Bardzo bogata bibliotekę i przyjemna składnie sprawia, że język Python jest bardzo często wyrkozystywany.

Jako ciekawostkę można powiedzieć, że nazwa języka Python wzięła się nie od zwierzęcia, lecz od słynnego Brytyjskiego serialu "Monty Python's Flying Circus"

Logo:

2.3.7 TensorFlow

Jest to biblioteka programistyczna wykorzystywana w uczeniu maszynowym i głębokich sieciach neuronowych. Została wydana jako otwarte oprogramowanie przez Google Brain Team w dniu 9 listopada 2015.

Umożliwia pisanie programów m. in. językach takich jak Python czy C. Jest dostępny na 64-cztero bitowych systemach operacyjnych: Windows, Linux, macOS oraz na platformach mobilnych: Android oraz iOS. Zaś w maju 2017 został wydany TensorFlow Lite jako dedykowane rozwiązanie specjalnie dla użytkowników Androida.

Olbrzymią zaletą TensorFlow jest to, że reprezentuje on paragdyman Dataflow, w którym program ma postać grafu skierowanego modelującego przepływ danych pomiędzy niezależnymi operacjami w węzłach. W pierwszym kroku należy zdefiniować model, a następnie stworzyć TensorFlow session

i ją uruchomwić. Takie podejście do pisania programów posiada ogromną zaletą jaką jest możliwość efektywnego programowania rozproszonego.

Tensorflow daje możliwość nie tylko wykorzystania CPU, ale równie dobrze korzystania z GPU. Wszystko to powoduje pełne wykorzystanie mocy obliczeniowej komputerów przy obliczeniach rozproszonych.

TensorFlow oprócz niskopoziomowych struktur posiada również moduły wyższego poziomu. Do modułów niskiego poziomu możemy odwoływać się poprzez wartwę API, która umożliwia łatwy do używania interfejs przeznaczony do używania w modelach błębokiego uczenia. Zać nad warstwą API znajduje się warstwa wysoko poziomowago API tj. Keras czy Estimator

2.3.8 Keras

Otwartoźródłowa biblioteka programistyczna napisana w języku Python wydana w dniu 27 marca 2015 napisana przez pierwotnego autora François Chollet.

Jej głównym przeznaczeniem jest to, aby w jak najprostszy i jak najszybszy sposób umożliwość pracę w głębokich sieciach nauronowych. Co więcej został zaprojektowana w sposób przyjemny do użytkowania, skupiona na modularności oraz rozszerzalności. Zaś od 2017 roku TensorFlow wspiera w swojej bibliotece Kerasa. Dzięki czemu Keras jest wysokopoziomową warstwą tworzenie modelów sieci neuronowych i jej uczenia.

2.4 Gra papier, kamień, nożyce

Rozdział 3

Wykonana praca

3.1 Aktywizacja danych

3.2 Detekcja ruchu

3.3 Detekcja kolorów

3.4 Szumy w zbiorze danych

3.5 Znajdowanie obiektów

3.6 Dobór cech

3.7 Ocena cech

3.8 Rozpoznawanie gestów przy pomocy klasyfikatora KNN

3.9 Rozpoznawanie gestów przy pomocy sieci neuronowej

3.9.1 Sieć neuronowa dla cech

3.9.2 Sieć neuronowa obrazów binarnych

Rozdział 4

Podsumowanie

Bibliografia