

Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie

Wydział Elektrotechniki, Automatyki, Informatyki i Inżynierii
Biomedycznej

KATEDRA BIOCYBERNETYKI I INŻYNIERII BIOMEDYCZNEJ



ROZPRAWA DOKTORSKA

Katarzyna Barczewska

Automatyczne rozpoznawanie gestów
na potrzeby wspomagania komunikacji z komputerem
osób z porażeniem mózgowym
ze szczególnym uwzględnieniem dzieci

Promotor:
prof. dr hab. inż.
Ryszard Tadeusiewicz

Kraków, 2018

Streszczenie

W związku z trudnościami w pozyskaniu danych od docelowych użytkowników, ograniczeniami, a także zróżnicowaniem ich indywidualnych predyspozycji, sprecyzowano szereg wymagań, jakie powinien spełniać system automatycznego rozpoznawania gestów dedykowany dla osób z porażeniem mózgowym. Na podstawie analizy sporządzonej listy wymagań sprecyzowano cel pracy. Celem było stworzenie systemu szybko uczącego się nowych gestów wykonywanych przez nowego użytkownika na podstawie niewielkiej liczby powtórzeń.

Dokonano przeglądu literatury poświęconej automatycznemu rozpoznawaniu gestów, począwszy od sposobów akwizycji danych, przez metody przetwarzania, po stosowane modele gestów. Na podstawie przeglądu opracowano koncepcję rozwiązania, w którym gesty dynamiczne, reprezentowane jako sekwencje podjednostek, modelowano przy pomocy równoległych ukrytych modeli Markowa (ang. *Parallel Hidden Markov Models*, PaHMM). Problem niewielkiej ilości danych gestów osób z porażeniem mózgowym rozwiązano inspirując się na metodzie transfer-learningu: do zbudowania modeli gestów osób niepełnosprawnych zaplanowano wykorzystać modele gestów osób zdrowych. Na podstawie zgromadzonej bazy gestów osób zdrowych wytrenowano modele podjednostek gestów (ukryte modele Markowa, HMM). Następnie wykorzystując leksykon wytrenowanych podjednostek wykonano ich najlepsze dopasowanie do gestów nowych osób. Najlepsze dopasowanie polegało na znalezieniu sekwencji podjednostek maksymalizującej kryterium największej wiarygodności wygenerowania danych obserwacji (algorytm Viterbiego). Dzięki wykorzystaniu stanów nieemitujących modeli podjednostek, modele podjednostek z rozpoznanej sekwencji „sklejono” w model nowego gestu, który był potem wykorzystywany przez system rozpoznawania. Do eksperymentów wykorzystano zarówno sensor Kinect 2.0 jak i czujnik inercyjny o 9 stopniach swobody. Eksperymenty przeprowadzono zarówno na grupie osób zdrowych jak i niepełnosprawnych. Wykonano też szereg dodatkowych analiz, w których wprowadzono nowy prosty wektor cech wystarczający do opisu gestykulacji, zaproponowano metodę automatycznego grupowania cech do kanałów modelu równoległego PaHMM, a także dokonano analizy doboru liczby kanałów. Opracowano metodę automatycznej indywidualnej rekomendacji gestów na potrzeby interakcji z komputerem, przy pomocy której można wskazać zestaw gestów, które najlepiej się sprawdzą w interakcji człowiek komputer. Zaproponowano również wykorzystanie większej liczby miar pozwalających na ocenę działania systemu automatycznego rozpoznawania gestów w warunkach rzeczywistego użytkowania.

Wykazano, że do stworzenia modeli, które pozwalają uzyskać wysokie wartości skuteczności rozpoznawania, wystarcza wykonanie przez nowego użytkownika chociaż jednego powtórzenia nowego gestu. Poruszana problematyka wpisuje się na listę najważniejszych kierunków rozwoju metod automatycznego rozpoznawania gestów. Zaproponowane rozwiązanie jest istotne z punktu widzenia rozwoju metod uczenia na podstawie małej liczby przykładów uczących, wykorzystania metod transfer-learningu, a także rozpoznawania gestów niezależnie od użytkownika. Wypracowane rozwiązania mogą mieć przełożenie na wiele innych zastosowań automatycznego rozpoznawania gestykulacji.

Abstract

Due to the difficulties in obtaining data from the users, limitations as well as the diversity of their individual predispositions, a number of requirements have been specified that should be met by the automatic gesture recognition system dedicated to people with cerebral palsy. Based on the analysis of the list of requirements, the goal of the work was specified. The goal was to create a system that would be able to learn new gestures made by an unknown user using only a small number of repetitions.

A review of the literature on automatic gestures recognition systems was made. The review included methods of data acquisition, data processing methods and different gesture models. Based on the review results, a solution concept was developed in which dynamic gestures, represented as sequences of subunits, were modeled using parallel hidden Markov Models (PaHMM). The problem of small amount of data of gestures performed by persons with cerebral palsy was solved by the transfer-learning method: models of healthy user's gestures were planned to be used to build models of the gestures of the disabled. On the basis of the collected database of gestures, models of gesture subunits were trained (hidden Markov models, HMM). Then, using the lexicon of the subunits, the best alignment of the subunits to the unknown gestures was performed. The best alignment consisted of decoding the sequence of subunits by the maximum likelihood criterion (Viterbi algorithm). Thanks to the use of non-emitting states of the subunit models, the subunits from the recognized sequence were "sticked" into a new gesture model, which was then used in the recognition system. Both the Kinect 2.0 sensor and the inertial sensor with 9 degrees of freedom were used in experiments. Experiments were carried out on both healthy and disabled people. A number of additional analyzes were carried out, in which a new simple feature vector was introduced, the method of automatic grouping of features to the PaHMM channels was proposed, and the selection of the number of channels was analyzed. A method of automatic individual recommendation of gestures for the purpose of interaction with a computer was developed. Automatic recommendation helps to indicate a set of gestures that works best in a human computer interaction. It was also proposed to use more evaluation metrics of the automatic gesture recognition system to assess its reliability in the real use.

It has been shown that to create models that allow to obtain high values of recognition accuracy, it is enough for the new user to make at least one repetition of a new gesture. The issues raised in this work are on the list of the most important directions for the development of automatic gestures recognition methods. The proposed solution is important for the development of learning methods which use a small number of training examples, the transfer-learning methods, and the recognition of gestures the user independent case. The developed solutions can be transferred into many other applications of automatic gesturing recognition.

Spis treści

1. Wstęp.....	5
1.1 Motywacja.....	5
1.2 Tezy, cele i zakres pracy	8
1.3 Układ pracy	9
2. Stan wiedzy	11
2.1 Wykorzystanie gestów osób z porażeniem mózgowym	11
2.2 Czym są gesty?	13
2.3 Akwizycja danych	14
2.4 Przetwarzanie danych	16
2.5 Automatyczne rozpoznawanie gestów	17
2.5.1 Scenariusze eksperymentów.....	17
2.5.2 Wykorzystywane metody.....	19
2.5.3 Sposoby adaptacji	20
2.5.4 One-shot-learning w rozpoznawaniu gestów	23
2.5.5 Zastosowania automatycznego rozpoznawania gestów.....	25
2.5.6 Wyzwania	25
3. Metody.....	28
3.1 Akwizycja i przetwarzanie danych	28
3.1.1 Dane z czujnika inercyjnego	28
3.1.2 Dane z sensora Kinect	29
3.1.3 Grupowanie cech do kanałów PaHMM.....	32
3.2 Modelowanie gestów.....	36
3.2.1 Ukryte modele Markowa	37
3.2.2 Trenowanie modeli 	43
3.2.3 Rozpoznawanie gestów.....	45
3.2.4 Równoległe ukryte modele Markowa	46
3.2.5 Fuzja modeli równoległych	46
3.3 Idea najlepszego dopasowania podjednostek	47
3.4 Ewaluacja systemu	52
3.4.1 Metody walidacji	52
3.4.2 Miary oceny działania systemu	53
4. Eksperymenty	57
4.1 Bazy danych.....	57

4.1.1	Baza WiTKOM:.....	58
4.1.2	Bazy IMU.....	59
4.1.3	Bazy Kinect	61
4.1.4	Baza DR-TEST	63
4.2	Scenariusze eksperymentów.....	65
5.	Dyskusja wyników.....	68
5.1	Wyniki dla baz IMU i IMU-szkoła.....	68
5.1.1	Alternatywa modeli dla bazy IMU-SZKOŁA	70
5.1.2	Rekomendacja zestawu gestów	73
5.1.3	Podsumowanie	81
5.2	Wyniki dla baz Kinect.....	82
5.2.1	Automatyczne grupowanie cech	82
5.2.2	Wyniki dla baz 017.....	85
5.2.3	Leksykony podjednostek	86
5.2.4	Wyniki dla bazy DR-TEST	88
5.2.5	Dobór kanałów modelu równoległego dla bazy DR-TEST	94
5.2.6	Alternatywa modeli dla bazy DR-TEST	103
5.2.7	Statystyki dla liczby podjednostek przypadających na gest.....	105
5.2.8	Podsumowanie.....	107
6.	Wnioski.....	109
7.	Podsumowanie	112
7.1	Oryginalne elementy pracy	113
7.2	Trudności	114
7.3	Możliwe kierunki rozwoju	114
	Załącznik 1: Eksperymenty z grami na tablet	116
	Załącznik 2: Aplikacja demonstracyjna.....	120
	Bibliografia	122

Skróty i oznaczenia wykorzystane w pracy:

ACC	skuteczność klasyfikatora (ang. <i>accuracy</i>);
ASLR	automatyczne rozpoznawanie języka migowego (ang. <i>automatic sign language recognition</i>);
CMLLR	(ang. <i>Constrained Maximum Likelihood Linear Regression</i>);
CPCC	(ang. <i>cophenetic correlation coefficient</i>);
Cov	macierz kowariancji;
CRF	warunkowe pola losowe (ang. <i>Conditional Random Fields</i>);
EER	błąd zrównoważony (ang. <i>equal error rate</i>);
EM	metoda największej wiarygodności (ang. <i>Expectation-Maximization</i>);
ERR	błąd rozpoznawania (ang. <i>error rate</i>);
FN	falszywie negatywny (ang. <i>false negatives</i>);
FNR	(ang. <i>False Negative Rate</i>);
FP	falszywie pozytywny (ang. <i>false positives</i>);
FPR	(ang. <i>False Positive Rate</i>);
GMM	Gaussowski model mieszany (ang. <i>Gaussian mixture model</i>);
HC	klasteryzacja hierarchiczna (ang. <i>hierarchical clustering</i>);
HC-PCA	metoda automatycznego grupowania cech do niezależnych kanałów PaHMM wykorzystująca klasteryzację hierarchiczną i analizę składowych głównych;
HCI	interakcja człowiek-komputer (ang. <i>human computer interaction</i>);
HMM	ukryte modele Markowa (ang. <i>Hidden Markov Models</i>);
HOG	histogramów zorientowanych gradientów (ang. <i>histogram of oriented gradients</i>);
HRI	interakcja człowiek-robot (ang. <i>human robot interaction</i>);
HTK	pakiet do trenowania modeli HMM (ang. <i>Hidden Markov Model Toolkit</i>);
IMU	czujnik inercyjny (ang. <i>inertial measurement unit</i>);
IMU 9DoF	czujnik inercyjny o 9 stopniach swobody;
KB	nazwa wektora cech zaproponowanego przez autorkę dysertacji;
MAP	(ang. <i>Maximum A Posteriori</i>);
MLLR	(ang. <i>Maximum Likelihood Linear Regression</i>);
MPD	mózgowe porażenie dziecięce;

NCBiR	Narodowe Centrum Badań i Rozwoju;
Nev	nazwa wektora cech zaproponowanego przez Neverową w pracy [1];
NN	sieć neuronowa (ang. <i>neural network</i>);
norm	prosty sposób przetwarzania wektora cech polegający na standaryzacji ich wartości;
PaHMM	równoległe ukryte modele Markowa (ang. <i>Parallel Hidden Markov Models</i>)
PaHMM_NN	równoległe modele Markowa z fuzją wykonaną za pomocą sieci neuronowej;
PaHMM_W	równoległe modele Markowa z fuzją ważoną;
PCA	analiza składowych głównych (ang. <i>Principal Components Analysis</i>);
Prec	precyzja klasyfikatora (ang. <i>precision</i>);
Recall	czułość klasyfikatora (ang. <i>sensitivity</i> lub <i>recall</i>);
ROC	krzywą ROC (ang. <i>Receiver Operating Curve</i>);
SVM	maszyna wektorów nośnych (ang. <i>Support Vector Machine</i>);
TN	prawdziwie negatywny (ang. <i>true negative</i>);
TP	prawdziwie pozytywny (ang. <i>true positive</i>);
UD	model zależny od użytkownika (ang. <i>user dependent</i>);
UI	model niezależny od użytkownika (ang. <i>user independent</i>);
VFH	deskryptor kształtu (ang. <i>Viewpoint Feature Histogram</i>);
WiTKoM	Wirtualny Tłumacz Komunikacji Migowej;
017-N	baza gestów wykonywanych w naturalny sposób;
017-NWS	baza gestów wykonywanych w naturalny, przesadnie delikatny i przesadnie dynamiczny sposób; każdy sposób wykonywania gestykulacji traktowany jako inny użytkownik ;
017-All-in-1	baza gestów wykonywanych w naturalny, przesadnie delikatny i przesadnie dynamiczny sposób; wszystkie sposoby gestykulacji składają się na pojedynczego użytkownika;