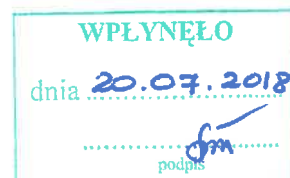




**Politechnika Łódzka**

Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej



dr hab. inż. Adam Wojciechowski  
Instytut Informatyki  
Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej  
Politechniki Łódzkiej

Łódź, 12.7.2018

### **RECENZJA ROZPRAWY DOKTORSKIEJ**

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

**Autor rozprawy:** mgr inż. Katarzyna Barczewska

**Promotor:** prof. dr hab. inż. Ryszard Tadeusiewicz

**Dziedzina:** Nauki techniczne

**Dyscyplina:** Biocybernetyka i inżynieria biomedyczna

#### **1. Jaki jest problem naukowy (teza) rozprawy i czy został on trafnie i jasno sformułowany?**

Problem rozpoznawania gestów jest jednym z najmocniej eksploatowanych obszarów biocybernetyki i inżynierii biomedycznej. Zainteresowanie naukowców tym obszarem jest dodatkowo stymulowane rolą jaką odgrywa gestykulacja w komunikacji niewerbalnej pomiędzy ludźmi i sprawność z jaką człowiek jest w stanie wykonywać semantyczne gesty, która z kolei przekłada się na wysoką pojemność informacyjną gestykulacji. Zastosowania metod rozpoznawania gestów rąk obejmują między innymi manipulację obiektami w środowiskach wirtualnych, interpretację znaków języka migowego, czy komunikację człowieka z robotami i komputerami, któremu to zagadnieniu Autorka rozprawy poświęciła najwięcej uwagi. Jednymi z beneficjentów opracowanych rozwiązań, w zakresie komunikacji człowiek-komputer, są osoby z porażeniem mózgowym, u których pogorszenie koordynacji wzrokowo ruchowej wpływa negatywnie na skuteczność wykonywania czynności życiowych (np. obsługi komputera) i tym samym obniża komfort życia.

Sformułowanie tezy pracy zostało poprzedzone bardzo szeroką analizą wymagań funkcjonalnych dzieci z porażeniem mózgowym i opiniami ich opiekunów. Podkreśliły one znaczną złożoność zadania, które postawiła sobie w rozprawie Autorka. Wśród najważniejszych wyzwań należy wymienić:

- unikalność gestów każdego z użytkowników, przy zachowaniu tolerancji w ramach poszczególnych klas gestów;
- konieczność szybkiej adaptacji (kalibracji) systemu rozpoznawania gestów do nowych użytkowników, połączona z wymogami wysokiej skuteczności ich rozpoznawania.

Brak uniwersalnych metod w zakresie modelowania gestykulacji w połączeniu z bardzo ograniczonym zestawem treningowym dla poszczególnych klas gestów, narzucił dodatkowe ograniczenia na rozwiązywany problem naukowy.

Ostatecznie sformułowana teza rozprawy brzmi dość ogólnie („Z wykorzystaniem metod sztucznej inteligencji można skonstruować system, który sam uczy się nowych gestów wykonywanych przez nowego użytkownika na podstawie niewielkiej ilości materiału treningowego”), ale doprecyzowuje ją pięć tez szczegółowych, które dogłębnie definiują

zadania i problemy postawione w rozprawie doktorskiej. Tym samym opisany problem naukowy został przez Autorkę dysertacji właściwie sformułowany i wyjaśniony.

Autorka właściwie definiuje pojęcie gestu, a skoncentrowanie się na gestach rąk nie ogranicza funkcjonalności i użyteczności zaproponowanych rozwiązań.

## **2. Czy tematyka rozprawy jest aktualna lub dostatecznie ważna?**

Tematyka rozprawy jest ważna i aktualna, szczególnie w odniesieniu do zastosowań w obszarze dynamicznie rozwijającego się przemysłu inteligentnych urządzeń oraz szeroko pojętej interakcji człowiek-komputer. Tematyka rozprawy ma również bardzo istotne znaczenie społeczne, w szczególności w odniesieniu do osób niepełnosprawnych ruchowo. Dysertacja wskazuje bezpośrednich beneficjentów opracowanych rozwiązań – dzieci z porażeniem mózgowym, które zachowując wysoką sprawność intelektualną borykają się z problemami obsługi urządzeń i komputerów, których wszechobecność w dzisiejszych czasach jest niezaprzeczalna. Osiągnięcia pracy wpisują się w proces usprawniania i zwiększania użyteczności komunikacji człowieka z otaczającym go światem, w szczególności osób dotkniętych porażeniem mózgowym, czego aktualność potwierdza szerokie zainteresowanie firm i liczne projekty naukowo-badawcze wspierane przez Narodowe Centrum Badań i Rozwoju.

Konsekwencją systematyzowania i usprawniania rozwiązań w obszarze naturalnych interfejsów użytkownika (ang. *natural user interfaces*) jest zwiększenie dostępności narzędzi, które znajdują swoje zastosowanie w obszarze medycyny, rozrywki, administracji publicznej, czy życia codziennym. Komfort życia i dostępność środowisk pracy dla osób z różnymi formami niepełnosprawności ruchowej, jest między innymi funkcją użyteczności i jakości systemów sterowania, ze szczególnym naciskiem na rozpoznawanie gestów.

Odpowiednie sprecyzowanie wymagań systemu, sprofilowane pod kątem osób niepełnosprawnych ruchowo, powoduje, że poruszana problematyka wpisuje się w listę najważniejszych kierunków rozwoju metod automatycznego rozpoznawania gestów. Przedstawione rozwiązanie jest istotne zarówno z punktu widzenia rozwoju metod uczenia na podstawie małej liczby przykładów uczących, wykorzystania metod *transfer-learningu*, a także rozpoznawania gestów niezależnie od użytkownika.

## **3. Czy autor rozwiązał postawiony problem i użył do tego właściwych metod?**

Osiągnięcie głównego celu pracy i weryfikacja postawionej w dysertacji tezy głównej, Autorka zdekomponowała na pięć cząstkowych hipotez badawczych, które brzmią następująco:

- *metodologię modelowania gestów języka migowego można wykorzystać na potrzeby modelowania innych gestów. Połączenie tej metodologii z elementami automatycznej analizy danych może poprawić jakość działania systemu;*
- *każdy nieznanany gest dynamiczny, niekoniecznie będący gestem języka migowego, można zakodować przy pomocy sekwencji znanych podjednostek gestu. Dekodowanie sekwencji podjednostek może się odbywać z wykorzystaniem kryterium najwyższej wiarygodności;*
- *model nieznanego gestu można zbudować poprzez połączenie znanych modeli podjednostek gestu odpowiadających odkodowanej sekwencji. Połączenie sekwencji modeli podjednostek w modelu gestu może zostać zrealizowane poprzez iloczyn prawdopodobieństw;*
- *do zbudowania modelu nowego gestu pozwalającego na uzyskanie wysokiej wartości skuteczności rozpoznawania wystarczy pojedyncze powtórzenie nowego gestu;*
- *stworzenie alternatywy modeli dla jednego gestu pozwoli lepiej odzwierciedlić osobnicze zróżnicowanie w sposobie wykonywania tego samego gestu, a tym samym wpłynie na wzrost skuteczności rozpoznawania;*

W obszarze ekstrakcji cech sygnałów, pochodzących z wykorzystywanych w badaniach sensorów (*IMU, Kinect 2.0*), Autorka przyjęła nieco rozbudowany, ale opisany i uznany w literaturze sposób reprezentacji danych pomiarowych: wielkości fizyczne zarejestrowane przez czujniki pomiarowe (prędkość kątowa, przyspieszenie), położenie punktów anatomicznych szkieletu postaci i ich odległość od głowy, kąty pomiędzy segmentami szkieletu postaci oraz kąty pomiędzy segmentami szkieletu, a osiami ciała. W przypadku kontrolera *Kinect* najlepsze wyniki osiągnięto dla prostego wektora cech zaproponowanego przez Autorkę - znormalizowanych wartości położenia, prędkości i przyspieszenia wybranych stawów szkieletu ręki i ich odległości od głowy.

W zakresie modelowania gestów dynamicznych Autorka wykorzystwała równoległy ukryty model Markowa (PaHMM), który wcześniej był głównie stosowany w rozpoznawaniu gestów języka migowego. Podejście wielokanałowe, w metodach uczenia maszynowego, jest uzasadnione znaczną liczbą części ciała, które biorą równocześnie udział w wykonywaniu gestu. Doświadczenia wykonane przez Autorkę rozprawy jednoznacznie wykazały przewagę modelu PaHMM nad zwykłym modelem HMM, pod względem skuteczności rozpoznawania gestów, przy równoczesnym obniżeniu wartości błędu zrównoważonego, istotnym z punktu widzenia niezawodności działania systemu. Istotny wpływ, na usprawnienie rozpoznawania gestów, miały wprowadzone przez Autorkę metody automatycznego grupowania cech spełniających założenia niezależności poszczególnych kanałów PaHMM i diagonalnej postaci macierzy kowariancji wewnątrz pojedynczego kanału PaHMM. Pierwsze założenie spełniono za pomocą grupowania cech dokonanych z wykorzystaniem procedury aglomeracyjnej klasteryzacji hierarchicznej, bazującej na macierzy odległości korelacji par cech. Drugie założenie zostało spełnione dzięki analizie głównych składowych cech (ang. *principal components analysis*).

Wyniki skuteczności rozpoznawania gestów, uzyskane przy zastosowaniu podejścia automatycznego grupowania cech do kanałów modelu równoległego nie poprawiły znacznie skuteczności klasyfikacji względem podejścia zgodnego z semantycznym grupowaniem cech gestów, ale znacznie usprawniły metodykę postępowania w sytuacjach, gdy podział cech nie jest jednoznaczny. Automatyzacja grupowania cech pozwala dodatkowo na walidację poszczególnych zestawów cech, których selekcja zapewnia nie tylko większą szybkość dostosowania się systemu do gestów nowego użytkownika, ale również pozwala na optymalizację liczby kanałów w modelu równoległym – badania wykazały, że można wskazać kombinację mniejszej liczby kanałów, które działają skuteczniej niż podejście wykorzystujące wszystkie grupy cech w ramach wszystkich kanałów modelu. Zwieńczeniem badań Autorki jest połączenie modelu PaHMM z automatycznym grupowaniem cech oraz selekcją kanałów, która daje zarówno najlepsze wyniki skuteczności rozpoznawania gestów, ale także istotnie upraszcza model ich reprezentacji. Pozwala to w efekcie na lepszą generalizację wiedzy oraz szybsze dostosowanie systemu do nowych użytkowników, co jest szczególnie istotne dla dzieci z porażeniem mózgowym.

Istotnym komponentem opracowanego rozwiązania problemu naukowego, decydującym bezpośrednio o skuteczności rozpoznawania gestów, był odpowiednio wygenerowany leksykon podjednostek ruchu/gestu, zwanych również cheremami. Odpowiednie ich zdefiniowanie, na podstawie danych treningowych, a następnie połączenie modeli podjednostek w modele gestów, zapewniły wysokie wartości skuteczności rozpoznawania gestów testowych. Badania wykazały, że każdy z wybranych gestów dynamicznych (baza DR-TEST) można było opisać za pomocą sekwencji zdefiniowanych wcześniej podjednostek, posługując się kryterium najwyższej wiarygodności i algorytmem Viterbiego. Badania z wykorzystaniem bazy gestów spoza zbioru treningowego, jak również gestów wykonywanych przez osoby z porażeniem mózgowym wykazały, że niewielka liczba (w skrajnym przypadku równa nawet 1) powtórzeń nowego gestu pozwala na zbudowanie jego modelu przy zachowaniu wysokiej skuteczności jego rozpoznawania. Przeprowadzone

zostały również eksperymenty w zakresie personalizacji modeli gestów za pomocą wybranych metod adaptacji modeli (m. in.: CMLLR, MAP).

Dodatkowa optymalizacja modelu wielokanałowego dokonana została w zakresie doboru liczby i rodzaju kanałów modelu opisującego gesty. Przeprowadzone eksperymenty z użyciem większej liczby powtórzeń nieznanego gestu, powodujące konstruowanie alternatywy modeli gestu, zdefiniowanej przez zdekodowane za każdym razem sekwencje podjednostek ruchu, doprowadziły do wniosków, że rozwiązanie to pozwala na zwiększenie skuteczności rozpoznawania gestów. Wykorzystanie alternatyw modeli gestu umożliwia opisanie większej różnorodności osobniczej w sposobie gestykulacji, co zwiększa przydatność opracowanego rozwiązania dla osób z porażeniem mózgowym, które to osoby mają trudności z idealnym powtarzaniem tych samych gestów. Z drugiej strony, eksperymenty wykazały, że nieumiejętne wykorzystanie nawet dużej bazy danych treningowych może prowadzić do gorszego efektu niż selektywne wykorzystanie małej bazy nagrań treningowych.

Bardzo ciekawym rozwiązaniem problemu niedostatków skuteczności rozpoznawania gestów osób niepełnosprawnych ruchowo była, zaproponowana przez Autorkę rozprawy, metoda rekomendacji gestów dla indywidualnego użytkownika, która dodatkowo zwiększyła użyteczność opracowanych rozwiązań.

W końcowej części pracy Autorka przedstawiła wyniki i dyskusję badań oraz eksperymentów, które stanowią niezwykle istotną wartość pracy. Zagadnienie rozpoznawania gestów, które z pozoru wydaje się mocno wyeksploatowane, dzięki odpowiednio zaplanowanym eksperymentom, odpowiedniej metodologii badawczej, błyskotliwemu wnioskowaniu oraz ogromnemu nakładowi pracy na weryfikację cząstkowych hipotez i wariantów, doprowadziło do bardzo ciekawych i interesujących wniosków, które za każdym razem zostały zaprezentowane z wykorzystaniem szerokiego spektrum miar i zostały udowodnione statystycznie.

Bardzo interesującymi eksperymentami były doświadczenia z ewaluacją różnych leksykonów podjednostek ruchu w kontekście najlepszego dopasowania podjednostek do jednego powtórzenia gestu, jak również eksperymenty z walidacją poszczególnych kanałów (grup cech) równoległego modelu gestu.

Przeprowadzone eksperymenty cząstkowe, zostały zrealizowane zgodnie z obowiązującą metodologią badawczą i pozwoliły na udowodnienie hipotez cząstkowych, które z kolei stanowią mocny i niepodważalny dowód pozytywnej weryfikacji tezy głównej dysertacji.

#### **4. Na czym polega oryginalny dorobek autora i jakie jest jego znaczenie poznawcze lub przydatność praktyczna dla nauki bądź techniki?**

Wgłębiając się w poszczególne etapy metodyki rozpoznawania gestów dynamicznych łatwo zauważyć, że Autorka rozprawy zaproponowała szereg cennych i wartościowych metod oraz usprawnień wpływających korzystnie na skuteczność rozpoznawania gestów rąk. W mojej ocenie na główne, oryginalne dokonania rozprawy doktorskiej składają się:

- opracowanie metody uczenia maszynowego, w miejsce stosowanych powszechnie metod opartych o reguły, do personalizacji modelu gestów rozpoznawanych przez system, na podstawie pojedynczych powtórzeń poszczególnych gestów;
- metoda budowania leksykonu podjednostek ruchu/gestu na podstawie gestów osób zdrowych, a następnie wykonanie ich najlepszego dopasowania (algorytmem Viterbiego) do gestów nowych osób;
- opracowanie alternatywy modeli gestu celem zwiększenia skuteczności ich rozpoznawania u osób niepełnosprawnych ruchowo;

- opracowanie metody automatycznego podziału cech do równoległych kanałów modelu PaHMM, której skutkiem jest redukcja wymiaru wektora cech przy zachowaniu niemalże identycznej skuteczności rozpoznawania gestów, co wpływa na efektywność trenowania modeli i sprawność zarządzania cechami, gdy ich podział nie jest intuicyjny;
- opracowanie metody automatycznej, indywidualnej rekomendacji gestów na potrzeby interakcji z komputerem;

Istotnym osiągnięciem są również obszerne i drobiazgowo badania weryfikujące skuteczność zaproponowanych metod i optymalizujące parametry poszczególnych rozwiązań. Autorka zaproponowała i wdrożyła kompleksową ewaluację rozwiązań za pomocą dużo większej niż standardowa liczby miar. Wśród badań, które zasługują na szczególną uwagę należy wymienić:

- analizę selekcji kanałów i tym samym wyboru grup cech, wykorzystywanych w fuzji modelu równoległego PaHMM, które uzyskują najwyższą skuteczność rozpoznawania i minimalizują błąd zrównoważony;
- analizę liczby podjednostek i budowania leksykonów podjednostek ruchu, pozwalających na efektywne budowanie modeli gestów;

## 5. Czy rozprawa świadczy o dostatecznej wiedzy autora i znajomości współczesnej literatury, z dyscypliny naukowej, której dotyczy?

Autorka zawarła w pracy przegląd wybranych metod związanych z procesem akwizycji i przetwarzania danych opisujących gesty, ze szczególnym uwzględnieniem danych pochodzących z czujników inercyjnych (IMU), jak również sensorów głębi (*Microsoft Kinect 2.0*). Dużo uwagi zostało poświęcone zagadnieniu automatycznego rozpoznawania gestów, zagadnieniu adaptacji istniejących modeli rozpoznawania gestów na potrzeby nowych użytkowników, w tym przy wykorzystaniu tzw. *transfer-learning*, jak również problemowi adaptacji modeli rozpoznawania na podstawie niewielkiej liczby powtórzeń gestów (tzw. *one-shot-learning*). Przegląd zrealizowany został systematycznie, według kolejnych etapów metodyki rozpoznawania gestów. Końcowym etapem przeglądu są wyzwania naukowe, które stanowią nie tylko kanwę dla badań Autorki, ale mogą również stanowić inspirację dla osób czytających rozważaną dysertację.

W kontekście opracowanych rozwiązań Autorka właściwie dobrała rozwiązania referencyjne, do których odnosiła się w swojej pracy, lub które modyfikowała wprowadzając istotne usprawnienia. Zaproponowane przez Autorkę metody i rozwiązania są oryginalne i zostały uznane poprzez kolegia redakcyjne, o czym świadczą publikacje w czasopiśmie i prezentacje na międzynarodowych konferencjach. Bibliografia zawiera 111 pozycji o uznanej renomie, właściwie dobranych, z czego 3 prace są współtworzonych przez Autorkę rozprawy (w tym jedna z listy JCR).

## 6. Jakie są wady i słabe strony rozprawy?

Praca nie zawiera żadnych istotnych uchybień, a poziom jej opracowania redakcyjnego i merytorycznego należy uznać za bardzo wysoki. Do drobnych uwag, niepodważających mojej bardzo pozytywnej opinii o całości rozprawy, chciałbym zaliczyć kilka kwestii szczegółowych:

- zamieszczony na str. 34 wzór 3.14 zdaje się być pozbawiony parametru (t) przy elementach wektora cech po prawej stronie równania;
- symbol  $Q'_T$  w pierwszym kroku algorytmu 3.4 (Viterbiego) powinien prawdopodobnie być zastąpiony symbolem  $Q'_t$ ;
- na rys. 4.1 zamieszczono schematy trajektorii gestów, które wyglądają na identyczne, a prawdopodobnie nie miały być takie same, np. gest 3 i 4 oraz gest 5 i 6;

- rysunek 4.6 przedstawia trajektorie ruchu dłoni dla zestawu gestów wymyślonych przez kolejne osoby. Pomocne byłoby umieszczenie ich symbolicznego schematu w sąsiedztwie znormalizowanych i zagręgowanych trajektorii;
- przykład aplikacji (gry), opisanej w załączniku 1 pracy, której elementem było sterowanie za pomocą gestów ręki, mógłby zostać jawnie osadzony w kontekście konkretnych gestów, które trzeba było wykonać na potrzeby tworzenia bazy danych IMU-SZKOŁA, podczas gdy to mapowanie nie jest do końca jednoznaczne;
- w pracy można znaleźć bardzo nieliczne błędy: literowe („techniki związane” - str. 11, „metody uch” - str. 11, „być też” - str. 17, „sa” - str. 38, „początkowej” - str. 38, „w wynikami” - str. 47, „dodatkowej personalizacja” - str. 58, „zakreślenia ręka” - str. 60, „każdego ze sposób” - str. 86, „jest też” - str. 92, „wzrost” - str. 100, „dla którym” - str. 103, „przyklejona niebieską taśma” - str. 117), edytorskie (brak numerów rysunków w tekście pracy, np.: „Rys.” i „Rys.()” - str. 50, „Rys.” - str. 118) i ortograficzne („nie zauważeniem” - str. 60), które w kontekście całej pracy są prawie niezauważalne i nie podważają bardzo wysokiej, całościowej oceny dysertacji.

Praca jest przygotowana bardzo starannie pod względem edycyjnym, z zamieszczonym spisem skrótów i symboli oraz spisem literatury, ułożonej chronologicznie, zgodnie z kolejnością pojawiania się poszczególnych pozycji w dysertacji.

## 7. Wniosek końcowy

Bardzo dobrze oceniam poziom merytoryczny rozprawy. Autorka wykazała się obszerną wiedzą w obszarze automatyzacji rozpoznawania dynamicznych gestów. Opracowała oryginalne metody oraz usprawnienia istniejących rozwiązań w zakresie przetwarzania i analizy cech gestów oraz modeli je opisujących.

Stwierdzam, że Autorka wniosła istotny wkład do dyscypliny naukowej biocebernetyka i inżynieria biomedyczna, w zakresie metod automatycznego rozpoznawania gestów dynamicznych.

Uważam, że przedłożona do recenzji rozprawa mgr inż. Katarzyny Barczewskiej spełnia wymagania Ustawy o stopniach naukowych i tytule naukowym z 14 marca 2003 roku, Dziennik Ustaw Nr 65, poz. 595 (z późniejszymi zmianami) odnośnie stopnia doktora nauk technicznych i wnoszę o dopuszczenie Autorki do dalszych etapów postępowania.

Wnioskuje również o jej wyróżnienie, z uwagi na bardzo dogłębną i drobiazgową analizę zagadnienia rozpoznawania dynamicznych gestów, zgodną z obowiązującą metodologią badawczą, zrealizowaną kompleksowo, według najwyższych standardów prowadzenia badań naukowych, które doprowadziły do wielu cennych wyników i płynących z nich wniosków. Na uwagę zasługuje również aktualność tematyki badawczej oraz jej istotna użyteczność społeczna.

Adam Wojciechowski