



PRZEWODNICZĄCY RADY DISCYPLINY  
Automatyka, Elektronika, Elektrotechnika  
i Technologie Kosmiczne  
prof. dr hab. inż. Wojciech Szelaż

Warszawa, 13.06.2023

dr hab. inż. Tomasz Trzeciński, prof. PW i UJ  
Instytut Informatyki Politechniki Warszawskiej  
Katedra Uczenia Maszynowego Uniwersytetu Jagiellońskiego

## RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY DISCYPLINY NAUKOWEJ AUTOMATYKA, ELEKTRONIKA I ELEKTROTECHNIKA POLITECHNIKI POZNAŃSKIEJ

**Tytuł rozprawy:** Combining interaction and perception to determine the physical properties of the robot environment

**Autor rozprawy:** Michał Bednarek

### 1 Analiza strony merytorycznej rozprawy

#### 1.1 Obszar problemowy

Recenzowana rozprawa doktorska dotyczy zagadnień związanych z percepcją robotyczną, w szczególności z wykorzystaniem sensorów haptycznych. Główną osią rozprawy jest rozwiązanie problemów percepcji przy wykorzystaniu współczesnych metod uczenia maszynowego, takich jak sztuczne sieci neuronowe. Aktualnie stosowane w robotyce metody polegają w dużej mierze na algorytmach wysokiego poziomu, których odpowiednie zaprogramowanie jest zadaniem inżynierów i konstruktorów. W związku z tym, że wraz z postępującą cyfryzacją mają oni dostęp do coraz większych zasobów danych, w tym danych zbieranych przez roboty w zadaniach percepcji otoczenia, naturalnym kierunkiem zainteresowań badawczych społeczności robotyków i automatyków jest uczenie maszynowe, a więc wykorzystanie danych do trenowania algorytmów kontroli robotów, a następnie testowanie ich w realistycznych okolicznościach, tj. poza starannie dobranymi warunkami laboratoryjnymi. Problem badawczy postawiony w pracy, jakim jest opracowanie metod bazujących na danych, które w lepszy sposób adaptują się do warunków rzeczywistych, jest więc nie tylko aktualny, ale wskazuje szereg zagadnień naukowych będących doskonałym polem badawczym i prowadzących do ważkich kontrybucji naukowych w dziedzinach takich jak informatyka czy robotyka.

Postawiony wyżej problem badawczy skłania Doktoranta do postawienia hipotezy, iż haptyczna percepcja może zostać osiągnięta przez roboty przy wykorzystaniu metod uczenia głębokiego, a więc bazujących na głębokich sztucznych sieciach neuronowych. Podążając w tym kierunku, niejako dopełniając postawioną tezę, Doktorant mapuje problem percepcji haptycznej jako zadanie translacji jednej sekwencji (danych pochodzących z sensorów) w drugą sekwencję (uzyskiwanej na temat środowiska informacji). Drugą tezę wspomagającą jest założenie, że połączenie wielu modalności rejestrowanych danych, w tym pochodzących zarówno z kamer jak i innych sensorów, w tym sensorów haptycznych, pomaga w zrozumieniu przez robota otaczającego go środowiska. Przedstawione tezy są możliwe do zweryfikowania empirycznego, na bazie zarówno istniejących jak i zaproponowanych przez Doktoranta zbiorów danych.

Celem pracy jest zbadanie i wdrożenie systemu składającego się z funkcjonalnych bloków, które umożliwiłyby i dodatkowo poprawiły ogólną jakość sensoryki haptycznej w robotyce. To określenie jasno i klarownie wskazuje obszar zainteresowań badawczych Doktoranta, jak również główną oś publikacji, wchodzących w skład rozprawy. Mieści się ono w pełni w dziedzinie robotyki oraz w takich działach informatyki technicznej jak widzenie maszynowe, uczenie maszynowe czy sztuczna inteligencja.

Rozprawa składa się z czterech części, podzielonych na rozdziały. Pierwsza część rozprawy to wstęp, który opisuje motywację podjęcia pracy badawczej w określonym wyżej kierunku, a także szerszy jej

kontekst. W części tej Autor opisuje pokrótce stawianą hipotezę badawczą, natomiast brakuje w nim moim zdaniem podsumowania wkładu w rozwój nauki przedstawionego w całej dysertacji. Część druga rozprawy składa się z trzech rozdziałów, które opisują proponowane przez autora metody do rozwiązania trzech problemów: rozpoznawania materiału na podstawie sygnału haptycznego oraz jego klasteryzacji, klasyfikacji terenu oraz połączenia wielu modalności. Każdy z tych rozdziałów zawiera wprowadzenie do problematyki, opis literatury przedmiotu oraz przedstawianie proponowanych metod. Następnie w części trzeciej rozprawy Autor przeprowadza weryfikację eksperymentalną zaproponowanych metod, również w podziale na trzy rozdziały, odpowiadające rozdziałom z części drugiej. Rozprawę kończy część czwarta, w której przedstawione zostały wnioski, również w podziale na podrozdziały odpowiadające poszczególnym zagadnieniom, a także uwagi końcowe.

Osobiście muszę przyznać, że zaproponowany przez Doktoranta podział znacząco utrudnia czytelnikowi skupienie nad tekstem i wyodrębnienie jasnej kontrybucji naukowej. Struktura ta generuje również potrzebę przeskoków pomiędzy sekcjami, jak np. na stronie 73 sekcja 5.2, która zawiera zdanie *I encourage readers to refer to the section 2.5 for formulating this topic*. Rozumiem motywację dotyczącą zwięzłego przedstawienia ewaluacji proponowanych metod, natomiast byłaby ona dużo silniejsza jeśli metody te składałyby się finalnie na jedno rozwiązanie. W przypadku opisywanych algorytmów są one jednak rozłączne i dotyczą innych problemów, co w połączeniu z rozbiciem na różne części pracy sprawia, że klarowność przekazu jest niska. Dodatkową komplikacją jest relatywnie wysokopoziomowe powiązanie poszczególnych tematów, od oceny sztywności do fuzji wielomodalnych sygnałów.

Podsumowując przedstawione w rozdziałach metody, można wyodrębnić następujące istotne problemy badawcze rozwiązywane przez autora:

1. wykorzystanie metod uczenia maszynowego do klasyfikacji materiału na bazie danych z sensorów haptycznych, zarówno wykorzystujące metody uczenia nadzorowanego, jak również metody uczenia nienadzorowanego,
2. opracowanie metody HAPTR bazującej na architekturze sieci neuronowej Transformer do klasyfikacji podłoża,
3. wykorzystanie metod łączenia modalności w zadaniach percepcji robotycznej.

Należy dodać, iż rozprawa zawiera reprezentatywną dla poruszanej problematyki bibliografię, która w większości przypadków właściwie ilustruje omawiane zagadnienia.

## 1.2 Ocena wyników oraz stopnia ich oryginalności

Recenzowana rozprawa dotyczy rozwiązania problemu percepcji robotycznej z wykorzystaniem zaczerpniętych z uczenia maszynowego metod i architektur. Autor wykorzystuje architektury głębokich sieci neuronowych takich jak spłotowe sieci neuronowe, rekurencyjne sieci neuronowe, czy wreszcie wspomnianą wyżej architekturę typu Transformer.

Zaprezentowane w rozprawie rozwiązania postawionych problemów badawczych bazują na istniejących rozwiązaniach i sprowadzają się w większości przypadków do połączenia ich w spójną całość, która rozwiązuje nietrywialny problem napotykanym w środowisku robotycznym. Ponad zaproponowane w rozprawie algorytmy, Autor przedstawia również sposób na wygenerowanie syntetycznego zbioru danych (sekcja 2.4.3), który służy do ewaluacji proponowanych metod.

Pomimo tego, iż część publikacji wchodzących w skład przedstawionej rozprawy ma kilku autorów, wkład Doktoranta nie jest wyraźnie opisany, choć należy domniemywać, że jako pierwszy autor większości z nich, miał on wkład znaczący. Niestety jednak w identyfikacji wkładu nie pomaga również sposób prezentacji, który w części drugiej rozprawy są wprowadzane zdaniami typu *Given the theoretical formulation above following could be introduced: [...]* (str. 33). Taka forma nie wskazuje jednoznacznie czy osobą odpowiedzialną za daną część pracy jest Autor czy inne osoby, czy wreszcie wynika ona naturalnie i jednoznacznie z poprzednich fragmentów pracy.

Pozostałe zauważone przeze mnie niedoskonałości dysertacji dotyczą przede wszystkim braku poprawnego umotywowania wyborów projektowych podczas tworzenia prezentowanych metod, jak również niepełnej ewaluacji i analizy uzyskanych wyników. Niedoskonałości te nie pozwalają w pełni zrozumieć metodyki projektowania opisanych rozwiązań, a przede wszystkim utrudniają identyfikację mocnych i słabych stron proponowanych metod. W poniższym zestawieniu zwracam uwagę na dostrzeżone niedostatki merytoryczne i proszę Doktoranta o ustosunkowanie się do nich.

- W rozdziale drugim, dotyczącym estymacji sztywności, pojawiają się architektury wykorzystujące sieci spłotowe oraz rekurencyjne. Brak jest wyników dla sieci w pełni połączonych (fully connected - FC). Proponowane rozwiązania zawierają oczywiście moduły FC, np. sieć RNN, ale nie dostrzegłem rozwiązań opierających się w całości na FC, co powinno być naturalnym punktem wyjścia rozważań. Czy Autor może przedstawić takie rozwiązanie i porównać jego wyniki z pozostałymi?
- W sekcji 5.2 przedstawione wyniki klasteryzacji pokazane są w formie wykresów słupkowych, bez oznaczenia wariancji uzyskanych rezultatów. Rezultaty uzyskiwane przez wiele z proponowanych algorytmów, np. k-means, zależy bezpośrednio od inicjalizacji punktów startowych, w związku z tym należałoby przedstawić rezultaty uzyskiwane na podstawie kilku prób, wraz z uwzględnieniem odchyłeń standardowych tych rezultatów. Czy Autor dysponuje takimi wynikami i mógłby je przedstawić?

### 1.3 Zagadnienia dyskusyjne

Poniższe uwagi dotyczą ogólniejszych kwestii poruszonych w rozprawie i nie odnoszą się bezpośrednio do treści pracy. Niemniej jednak liczę na analizę tych zagadnień i odniesienie się do nich przez Doktoranta.

- **Interpretowalność metody HAPTR przy wykorzystaniu modułu uwagi z Transformera:** ciekawym aspektem metody HAPTR zaproponowanej w rozdziale dotyczącym klasyfikacji podłoża jest wykorzystanie architektury Transformer oraz jej immanentnego elementu, tj. modułu uwagi. Ciekawym aspektem badań byłaby analiza tego na jakie elementy sygnału zwraca uwagę tenże mechanizm. Takie podejście mogłoby zwiększyć interpretowalność modelu i przybliżyć jego działanie. Czy Doktorant podejmował może próby takich eksperymentów?
- **Istnienie wiodącej modalności:** w rozdziale poświęconym ewaluacji metod fuzji różnych modalności, Autor pisze, że: *All experiments on robustness revealed the existence of a leading modality, which means that one modality always played a dominant role in the discrimination between classes.* To bardzo ciekawe sformułowanie, które mogłoby nasunąć wiele pomysłów jak taką wiodącą modalność wykorzystać, np. wykorzystując mechanizm ważenia odpowiednich części wektora wejściowego podczas treningu. Czy Doktorant podejmował takie próby modyfikacji algorytmów?
- **Brakująca pogłębiona analiza rezultatów dotyczących estymacji sztywności:** Na podstawie wyników prezentowanych w Tabeli 5.2 Autor omawia interesujące zjawisko, jakoby dokładność modelu rosła dla dużych wartości sztywności pudełek. Czy Doktorant może przedstawić hipotezy dotyczące wyjaśnienia tego zjawiska? Ponadto, skąd wynikały różnice pomiędzy pudełkami a obiektami w kształcie kuli?

## 2 Analiza strony formalnej rozprawy

### 2.1 Ocena układu pracy i redakcji manuskryptu

Recenzowana rozprawa doktorska napisana jest w języku angielskim i obejmuje w kolejności: streszczenie w języku angielskim, streszczenie w języku polskim, podziękowania, listę użytych skrótów oraz symboli, cztery części zasadnicze, a także bibliografię. Praca liczy w sumie 127 stron.

Bibliografia liczy 156 pozycje. W rozprawie wskazanych jest również 15 prac własnych autora, w tym 14 własnych. Prace te pochodzą z materiałów konferencyjnych jak również z czasopism branżowych. Spośród znajdujących się w bibliografii prac większość ukazała się w przeciągu ostatnich 10 lat co świadczy o dobrym umiejscowieniu tematyki rozprawy w aktualnym w skali światowej nurcie badań. Bibliografia rozprawy nie budzi zastrzeżeń od strony merytorycznej, a jej redakcja jest staranna i nie dostrzegłem w niej żadnych błędów.

Układ rozprawy został już przeze mnie omówiony, natomiast należy zaznaczyć, że sama praca jest bardzo starannie opracowana pod względem edytorskim i nie widać w niej żadnych błędów w sztuce. Użyta terminologia jest właściwa dla obszaru problemowego rozprawy, zarówno w zakresie robotyki, informatyki jak i uczenia maszynowego.

## 2.2 Uwagi szczególne

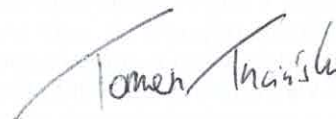
Tekst rozprawy jest w przeważającej większości poprawny pod względem językowym i stylistycznym. Podczas lektury zauważyłem nieliczne błędy redakcyjne i pomyłki wymagające korekty, np:

- str. 25, w. 6: "Most state-of-the-art recognition systems basis on artificial neural networks" - sugerowałbym zastąpić *basis* sformułowanie *is based on*, ponieważ lepiej pasuje ono do kontekstu jako czasownik.
- str. 25, w. 16-17: "However, a lack of interpretability in robotics might be a troublesome issue. However [...] such solutions still need to be explored" - sugeruję przepisanie, żeby uniknąć powtórzeń.
- str. 31, paragrafy "Feature Extractor" i "Recurrent Block" - rekomenduję uspoźnienie czasów, to przykład fragmentu, w którym następują po sobie dwie sekcje z różnymi czasami używanymi do opisu: w pierwszej jest to czas prosty teraźniejszy (*This module is responsible*), a w drugiej przeszły (*It processed [...]*). Sugeruję konsekwentne używanie czasu teraźniejszego.
- str. 42, w. 10: "MAL" - ten skrót został wyjaśniony tylko raz we wprowadzeniu. Warto go rozwinąć w tym rozdziale również.
- str. 93, w. 26: "No fixed level of noise can be considered universally unacceptable for a normalized input signal" - rekomendowałbym unikanie podwójnych zaprzeczeń, zaciemniając przekaz.
- str. 103, w. 8: "Considering that these experiments were targeted to prove or refute the existence of five tactile dimensions, one could summarize that they denied them, as there should be five of them as well." - brak klarowności, sugeruję uproszczenie tego zdania, ponieważ aktualny zapis utrudnia jego zrozumienie.
- str. 108, w. 1: "Throughout my Ph.D., I have significantly contributed to robotics and artificial intelligence, aiming at haptic perception for manipulators and walking robots." - rekomendowałbym bardziej powściągliwe używanie takich sformułowań. Ocena wkładu badacza w rozwój nauki powinna być wypadkową opinii innych badaczy oraz żywotności opublikowanych przez nią czy niego prac. Samodzielna ocena jest z natury rzeczy mało obiektywna.

## 3 Konkluzja

Uważam, że recenzowana dysertacja Pana mgr. inż. Michała Bednarka **spełnia wymogi** stawiane rozprawom doktorskim przez *Ustawę o stopniach naukowych i tytule naukowym z 14 marca 2003 roku (Dz. U. Nr 65 z 2003 r.)*, ponieważ zawiera rozwiązania istotnych problemów w dziedzinach informatyki i robotyki. Rezultaty nawiązują do aktualnego stanu wiedzy i mają znaczenie praktyczne, czego potwierdzeniem są liczne publikacje w czasopismach takich jak *Electronics* czy na renomowanych konferencjach (ICRA, ECMR, RoboSoft). Uzyskane wyniki eksperymentalne w wystarczającym stopniu dokumentują poprawność proponowanych koncepcji oraz skuteczność działania ich implementacji. Uwagi krytyczne sformułowane w treści recenzji, po części mające charakter dyskusyjny, nie wpływają istotnie na moją pozytywną ocenę poziomu merytorycznego przedstawionej rozprawy.

Sformułowanie problemu badawczego, sposób przeprowadzenia badań oraz zademonstrowana umiejętność formułowania wniosków świadczą o przygotowaniu Doktoranta do pracy naukowej. Na tej podstawie **wnioskuję o dopuszczenie mgr. inż. Michała Bednarka do dalszych kroków procedury uzyskania stopnia doktora nauk technicznych.**



Tomasz Trzcinski