

WPŁYNĘŁO
2025 -11- 27
dn.....

Warszawa, 11 listopada 2025

Prof. dr hab. inż. Janusz Kacprzyk, dr h.c. mult.
Członek rzeczywisty PAN
Fellow of IEEE (Life), IET, EurAI, AAIA, AIIA, IFIP, IAITQM, IFSA, SMIA
Instytut Badań Systemowych
Polska Akademii Nauk
Ul. Newelska 6
01-447 Warszawa
Email: kacprzyk@ibspan.waw.pl

Recenzja rozprawy doktorskiej mgr inż. Ekmeyong Awong Lisiane Esther pt. „Predicting human actions for utilisation in assistive robots”

Niniejsza recenzja została przygotowana w odpowiedzi na Uchwałę Nr. 155/III/2025 r. Rady Naukowej Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne” z dnia 30 września 2025 r., podpisane przez Przewodniczącego Rady, prof. dr. hab. inż. Pawła Szczepańskiego. Na mocy tej Uchwały zostałem wyznaczony jako jeden z recenzentów rozprawy doktorskiej mgr inż. Ekmeyong Awong Lisiane Esther pt. „Predicting human actions for utilisation in assistive robots”.

Przy przygotowywaniu recenzji wezmę przede wszystkim pod uwagę odpowiednie wymagania ustawowe dotyczące recenzji rozpraw doktorskich, wyrażone w art. 187, ust. 1 i 2 Ustawy Prawo o Szkolnictwie Wyższym i Nauce z dn. 20 lipca 2020 r.

Ponadto, ustosunkuję się także do następujących kwestii, zgodnie z dostarczonych mi zaleceniach Rady Dyscypliny, czyli:

1. Jakie zagadnienie naukowe/badawcze jest rozpatrywane w pracy (cel i teza rozprawy) i czy zostało ono jasno sformułowane przez autora?
2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł, w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle?
3. Czy autor rozwiązał postawione zagadnienie, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?
4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jak jest pozycja rozprawy w stosunku do stanu wiedzy i poziomu techniki reprezentowanych przez literaturę światową?
5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?
6. Jaka jest przydatność rozprawy dla nauk inżynieryjno-technicznych?

Na początek, chciałbym stwierdzić – o co coraz częściej także prosi się recenzentów – że praca jest prawidłowo zakwalifikowana do dziedziny nauk inżynieryjno-technicznych do Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne.

Ponadto, chciałbym też stwierdzić, że nie zachodzą żadne okoliczności typu konfliktu interesów, które mogłyby wywoływać jakiekolwiek wątpliwości co do mojej bezstronności przy sporządzaniu niniejszej recenzji.

Pierwszym pytaniem, jakie trzeba sobie zadać przy recenzowaniu rozpraw doktorskich, jest: czego praca dotyczy. W naszym przypadku recenzowana rozprawa dotyczy ogólnie jednego z najbardziej burzliwie się rozwijających kierunków współczesnej robotyki, jakim są tzw. roboty wspomagające zwane też robotami asystującymi. Są to roboty pomagające ludziom, zwłaszcza niepełnosprawnym, ale ogólniej wszystkim, którzy potrzebują pomocy, np. osobom w podeszłym wieku, wykonywać różne czynności, lub pomagać w wykonywaniu takich czynności, które stwarzają problemy tym osobom.

Nietrudno zauważyć, że ten dział robotyki rozwija się burzliwie, odpowiadając na rzeczywiste potrzeby nowoczesnych społeczeństw, które się starzeją w praktycznie wszystkich krajach rozwiniętych, a także ze względu na to, że takie roboty mogą do pewnego stopnia pomóc zaktywizować zawodowo duże grupy ludzi nie występujących dotąd na rynku pracy, co ma też wielkie znaczenie społeczno-ekonomiczne, zwłaszcza w przypadku braku siły roboczej.

Roboty wspomagające muszą jednak współdziałać z ludźmi, którzy są jednak w takim układzie oczywiście elementem dominującym. Można więc takie systemy „człowiek/ludzie + roboty wspomagające” uważać za świetny przykład tzw. „human centric system”, w którym człowiek odgrywa kluczową rolę. Niestety, ze względu na oczywiste różnice między człowiekiem i „maszyną” (tu robotem wspomagającym) wszelkie analizy, projektowanie, implementacja itp. takich systemów „human centric” są bardzo trudne i stanowią wyzwanie teoretyczne i techniczne.

A dalej, w tej wielkiej grupie zagadnień związanych ze współpracą człowieka z robotem wspomagającym, co właściwie powinno stanowić integrację, wśród różnych istotnych czynności, jedną z ważniejszych jest konieczność uzyskania przez projektowane systemy zdolności rozumienia i przewidywania ludzkich działań w czasie rzeczywistym, a tu kluczowym elementem jest umiejętność rozpoznawania postur ciała oraz wnioskowania o realizowanych działaniach, co umożliwi robotom proaktywne i intuicyjne reagowanie na zdarzenia. Jest to jednak wyjątkowo trudne ze względu na osobnicze różnice w wykonywaniu tych samych czynności i konieczność ekstrakcji niezmienniczych cech w zróżnicowanych warunkach, w jakich robot działa. Ponadto, ważna jest efektywność obliczeniowa i odporność systemu na zmienne warunki otoczenia, zmiany zachowania, percepcji itp. człowieka.

A zatem, można stwierdzić, że recenzowana praca doktorska:

- dotyczy zagadnienia o wielkim znaczeniu społeczno-ekonomicznym, ważnego dla praktycznie wszystkich krajów rozwiniętych,
- dotyczy zagadnienia trudnego z punktu widzenia analitycznego, algorytmicznego i implementacyjnego.

Jednocześnie, Kandydatka stosuje do analizy i rozwiązywania zadań bardzo nowoczesne metody oparte na danych (ang. data driven), w szczególności metody z wykorzystaniem zarówno nadzorowanych, jak i nienadzorowanych modeli uczenia maszynowego, stosując nowoczesną architekturę sieci neuronowej typu Transformer, co jest też bardzo nowoczesnym rozwiązaniem. Zastosowała także efektywny mechanizm wnioskowania inspirowany drzewami decyzyjnymi, co też jest dobrym rozwiązaniem. Ponadto, praca jest doskonałym

przykładem nowoczesnych tendencji z nauce i technice, ponieważ zawiera elementy wielu nauk, m.in. automatyki i robotyki, informatyki, wraz ze sztuczną inteligencją, nauki o danych itp.

Do implementacji opracowanych modeli, Kandydatka zaprojektowała i zbudowała autorski system z klasy HAP (Human Action Prediction), oparty na sekwencyjnym rozpoznawaniu postur ciała wyodrębnionych z danych RGB-D, a potem gruntownie go przetestowała na różnych referencyjnych zbiorach danych RGB-D obejmujące sceny o wysokim zagęszczeniu obiektów, dla różnych lokalizacji kamer.

Nie mam więc wątpliwości, że Kandydatka rozwiązała postawione zadanie we właściwy sposób, stosując właściwe metody i algorytmy, w dużej części będące Jej oryginalnym dorobkiem.

Trzeba tu od razu podkreślić, że wszystkie te powyższe cele cząstkowe, były projektowane, realizowane, implementowane i testowane przez Kandydatkę w bardzo obszerny i całościowy sposób, co czyni recenzowane rozprawę kompletną.

Nie mam więc wątpliwości, że:

- Rozpatrywany w pracy problem należy do nauk inżynierijno-technicznych, do dyscypliny automatyka, elektronika, elektrotechnika i technologie kosmiczne, choć oczywiście – co jest charakterystyczne dla wszystkich praktycznie nowych technologii – także inne dyscypliny, jak np. informatyka techniczna i telekomunikacja, mają tu duży wpływ,
- Problem jest ważny, ponieważ zawiera wyzwania badawcze, teoretyczne, związane z koniecznością zbudowania efektywnych algorytmów korekcji,
- Problem zawiera też wyzwania implementacyjne, w najlepszym duchu i tradycji nauk inżynierijno-technicznych,
- Problem ma też wielkie znaczenie ekonomiczne i społeczne.

Nie mam więc żadnych wątpliwości co do właściwego zaklasyfikowania doktoratu do odpowiedniej dyscypliny, a także do trafnego wyboru tematyki i zakresu.

Przejdę teraz do omówienia i oceny kolejnych rozdziałów i podrozdziałów pracy, co pozwoli na najbardziej właściwą ocenę rozprawy i jej niewątpliwych wielkich zalet.

Pierwszy rozdział, „Chapter 1: Introduction”, stanowi bardzo dobry przegląd głównych problemów i wyzwań w szeroko rozumianym obszarze robotów współpracujących z człowiekiem przy realizacji wielu różnych zadań, a zwłaszcza różnych aspektów dotyczących przewidywania działań człowieka w celu pogłębienia interakcji i jej efektywności. Omawia się wkład w tym zakresie innych nauk, jak np. nauk kognitywnych, psychologii, lingwistyki itp., ale też wizji komputerowej, nauk o danych itp.

Bardzo ciekawe są uwagi na temat rozpoznawania akcji ludzkich na podstawie informacji werbalnych, jak np. charakterystyka mowy czy informacje w języku naturalnym, i niewerbalnych, jak np. wyrazy twarzy, język ciała, postura ciała, gestykulacja itp.

Cenne są uwagi na temat ważności systemów z klasy HAP (Human Action Prediction), które mogą przewidywać ludzkie akcje przed ich pełnym wykonaniem. To ważny fragment pozwalający lepiej zrozumieć i docenić autorski system z tej klasy opracowany przez Kandydatkę dla potrzeb rozprawy. Uwagi na temat postur i tego, jak ich sekwencje tworzą ludzkie akcje, są też bardzo cenne i wyjaśniają wiele elementów pracy. Uwagi o danych trójwymiarowych są też bardzo ciekawe.

W dalszej części Kandydatka omawia znaczenie różnych metod uczenia maszynowego, m.in. nienadzorowanego i nadzorowanego oraz konieczność opracowania efektywnych algorytmów i ich implementacji. Ma to niewątpliwie związek z pracą.

Ciekawy jest też przegląd zastosowań, co może dostarczyć ciekawych inspiracji dotyczących tego, jakie są źródła danych i jakie metody analizy danych są najbardziej uzasadnione. Jednym z ciekawszych fragmentów w tym zakresie jest opis poświęcony pojazdom autonomicznym, ponieważ w tym przypadku występuje bardzo silna konieczność efektywnej współpracy pojazdu przede wszystkim z kierowcą, ale też z pieszymi, rowerzystami itp.

Podsumowując, moja ocena Rozdziału 1 jest bardzo pozytywna, Kandydatka wykazała się bardzo dobrą znajomością dziedziny robotów wspomagających, nawet w szerszym zakresie i kontekście współpracy człowieka z robotem, może nawet jeszcze szerzej w bardziej ogólnym zakresie „human centric systems”.

Może, choć to raczej byłby komentarz do przyszłych zamierzeń badawczych, dobrze było wspomnieć np. o zastosowaniu elementów tzw. obliczeń afektywnych, np. w ujęciu prof. Rosalindy Picard z MIT, do robotów wspomagających, zarówno indywidualnych jak społecznych. Być może jest to obiecujące rozwiązanie przyszłościowe.

Drugi rozdział, „Chapter 2: State-of-the-art”, zawiera w zasadzie bardzo obszerny i dokładny przegląd idei HAP (human action prediction), która jest szybko rozwijającym się działem sztucznej inteligencji zajmującym się rozpoznawaniem i przewidywaniem ludzkich akcji w czasie rzeczywistym. Jest to też często rozumiane jako klasa systemów implementujących metody i algorytmy te czynności realizujące. Nietrudno zauważyć, że kluczowym narzędziem pozwalającym na efektywną realizację HAP jest uczenie maszynowe i to jest rozpatrywane głównie w tym rozdziale. Przedstawia się bardzo dobrze analizę porównawczą stosowanych metod, i związanych z nimi algorytmów, pod kątem dokładności, efektywności numerycznej itp. Bardzo ciekawe i jednocześnie cenne jest to dodanie uwag o sensorach, z których pobiera się dane w konkretnym zastosowaniu, oraz jakie metody analizy danych są polecane. Uwagi o danych w postaci obrazów RGB, sygnałów RFID, sygnałów Wi-Fi, danych z sensorów przenośnych itp. są bardzo cenne i dobrze przedstawiają nowoczesne tendencje. Opis trudności i wyzwań stojących przed HAP, podany na stronie 38 rozprawy, jest bardzo dobrze zrobiony i bardzo użyteczny.

Cennym elementem pracy są też fragmenty poświęcone testowym zbiorom danych i benchmarkom, bez których trudne jest testowanie i ocena stosowanych algorytmów i tworzonych systemów.

Oprócz bardzo dobrze zrobionej analizy bardziej tradycyjnych metod analizy danych i uczenia maszynowego, jak np. metod probabilistycznych, maszyn wektorów podpierających, drzew decyzyjnych itp., bardzo cennym elementem jest tu krótki przegląd możliwości, jakie

oferują nowe rozwiązania z szeroko rozumianej klasy głębokiego uczenia, jak np. metody oparte na (głębokich) sieciach neuronowych, LLM (Large Language Models) czy na architekturze Transformer. To bardzo ciekawy fragment, a Tabela 2.2 zawierająca podsumowanie dokładności uzyskiwanych dla różnych omawianych metod jest bardzo cenna.

Podsumowując, Rozdział 2 stanowi bardzo wartościowe, choć zwięzłe, przedstawienie tego, co można znaleźć w zakresie nowych metod z szeroko rozumianego uczenia maszynowego w zastosowaniu do HAP. Jest bardzo dobrze zrobiony. Stanowi świetny punkt wyjścia dla nowych propozycji zaproponowanych w rozprawie.

Trzeci rozdział, „Chapter 3: Research Objective and Thesis Formulation”, jest poświęcony przedstawieniu celu i tezy rozprawy, ale pokazuje to w szerszej perspektywie, najpierw przedstawiając istotę zaproponowanego podejścia do HAP opartego na rozpoznawaniu ludzkich postur na podstawie danych RGB-D, koncentrując się na sekwencjach postur jako podstawowego elementu identyfikacji ludzkich akcji. W przeciwieństwie do metod tradycyjnych opartych na ruchu czy kontekście, praca koncentruje się na rozpoznawaniu różnych postur, które składają się na akcje. W tym kontekście, rozpatruje się różne nadzorowane i nienadzorowane metody rozpoznawania postur, analizuje się ich silne i słabe strony, a przede wszystkim podkreśla się konieczność stosowania rozwiązań bardziej elastycznych, efektywnych numerycznie oraz adaptowalnych do zmiennych warunków. W tym zakresie wskazuje się na pewne krytyczne braki w istniejącej metodologii i jej realizacji, m.in. niedostateczną możliwość predykcji długoterminowej, niedostateczną efektywność w przypadku bardzo dużych danych, niedostateczną możliwość analizy złożonych zadań itp.

W dalszej części rozdziału Kandydatka formułuje tezę pracy jako:

„Stosując rozpoznawanie postury ludzkiej dla predykcji akcji odbywającej się na bieżąco w środowisku opartym na danych, można uzyskać zadowalającą efektywność dotyczącą głównych wyzwań w HAP, co w końcu umożliwi otrzymanie bardziej adaptacyjnych i praktycznych systemów wspomagających człowieka”

przy czym przez zadowalającą efektywność rozumie się realizację operacji w czasie rzeczywistym w warunkach rzeczywistych z odpowiednią dokładnością i możliwością generalizacji na inne przypadki.

Ponadto, Kandydatka – po sformułowaniu powyższej tezy pracy – podaje dalej jako cele badawcze, które realizować będzie w następnych rozdziałach:

- opracowanie zmodyfikowanej architektury typu Transformer dla nadzorowanej klasyfikacji postur poprzez wprowadzenie nowych pojęć typu centralnej uwagi postury do uwagi transformacji ciała, które są wyznaczane stosując różne zbiory danych z użyciem standardowych wskaźników jakości (dokładność, precyzja, czułość, wskaźnik F1, itp.),
- opracowanie nienadzorowanego algorytmu klasteryzacji postur, który porówna się z zaproponowanymi modelami nadzorowanymi,
- wprowadzenie nowego wskaźnika, tzw. "Discriminant Score (DS)", do wspomagania wyboru optymalnego klastra i oceny jakości klasteryzacji postur w niezaetykietowanych zbiorach danych,

- zaproponowane metody ekstrakcji cech niezmienniczych wobec widoku, zwiększając odporność rozpoznawania postur dla różnych kątów ustawienia kamery i morfologii ciała,
- wprowadzenie mechanizmu rozumowania opartego na skalowalnej strukturze typu drzewa decyzyjnego reprezentującej ciągi postur i akcji jako elementy bazy danych, które zapewnią łatwe rozszerzania i aktualizację.

Te cele badawcze pokrywają praktycznie wszystkie ważne aspekty rozpatrywanego w pracy zadania naukowego, a także zadania projektowego, czyli opracowania autorskiej wersji systemu HAP. Kandydatka zdaje sobie sprawę, że choć większość prac dotyczących predykcji ludzkich akcji dotyczy ich działania w środowiskach laboratoryjnych, to przejście do analizy w warunkach rzeczywistych implikuje bez wątpienia konieczność zmiany pewnych założeń dotyczących np. jakości danych sensorycznych i konieczności ich wstępnej obróbki, zmienności środowiska, łagodnego przejścia między posturami, niezakłóconych potoków danych, stałości sposobów działania ludzi itp.

W pracy przyjmuje się założenia upraszczające, m.in. predykcje bez uwzględnienia kontekstu, brak konieczności dysponowania bardzo dużymi zbiorami danych, brak konieczności użycia skomplikowanych architektur i wielkich możliwości obliczeniowych, użycie prostych i efektywnych metod uczenia. Są one bardzo uzasadnione.

Biorąc pod uwagę powyższe założenia upraszczające, Kandydatka proponuje bardzo dobre rozwiązanie rozpatrywanego zadania predykcji akcji ludzkich na podstawie sekwencji postur z obrazów RGB-D. Zakłada prostą architekturę systemu HAP, które zawiera jako główne moduły:

- Akwizycji danych, ich czyszczenia i przetwarzania wstępnego,
- Rozpoznawania postur,
- Predykcja ludzkich akcji na podstawie rozpoznanych postur.

W pracy pokazuje się, że jest to bardzo efektywne rozwiązanie, z czym niewątpliwie należy się zgodzić.

A zatem, teza i cel rozprawy zostały sformułowane przez Kandydatkę w sposób jasny i dotyczą, z jednej strony, ważnego zagadnienia, stanowiącego jednocześnie wyzwanie, a z drugiej strony, tego, co Kandydatka rzeczywiście bardzo dobrze zrealizowała.

Rozdział czwarty, „Chapter 4: Data Acquisition and Preprocessing Techniques”, opisuje wstępny, choć niezwykle ważny etap realizacji postawionych w pracy zadań, czyli sposoby pozyskania danych i ich przygotowanie do użycia w procesie analizy, algorytmizacji i implementacji zaproponowanych w pracy rozwiązań. Punktem wyjścia jest tu proces predykcji ludzkich akcji, który standardowo obejmuje: (1) zebranie danych i ich wstępne przetwarzanie, co obejmuje m.in. lokalizację obiektu zainteresowania, tu ciała ludzkiego, śledzenie ciała ludzkiego, ekstrakcję sylwetki ciała ludzkiego, przetwarzanie danych, segmentację i odsumianie, (2) opracowanie reprezentacji danych, tzn. transformacja danych na format bardziej przystosowany do przetwarzania przez metodę uczenia maszynowego, ale zachowując ważne detale, stosując normalizacje itp., (3) opracowanie, trenowanie i testowanie wybranych algorytmów uczenia maszynowego oceniane z użyciem ustalonych metryk.

W tym rozdziale przedstawione bardzo dobrze zaproponowane rozwiązanie zbierania danych i ich przetwarzania wstępnego. Zastosowano rozwiązania, które mogą być obiecujące dla opracowywanego systemu typu HAP, czyli m.in. zawierające nowoczesne sensory RGB-D, ZED 2 i Microsoft Azure Kinect, które pozwoliły uzyskać dobrą jakość danych dotyczących RGB, głębi oraz szkieletów 3D. Pozwoliło to ominąć ograniczenia wizyjne tradycyjnych metod dwuwymiarowych. To dobre rozwiązanie i Kandydatce udało się uzyskać dobre i oryginalne wyniki. Jeśli chodzi o przetwarzanie wstępne, to celem jego stosowania było możliwie duże zmniejszenie różnic w danych zależnych m.in. od wielkości ciała, odległości kamery, kąta skierowania kamery, szybkości ruchu, różnic osobniczych itp., co może wprowadzać znaczny szum w danych. Przetwarzanie wstępne zawiera annotację i segmentację nagrań wideo dotyczących przejść w posturach, a nie tradycyjnych różnic w klatkach w czasie, co jest oryginalnym wkładem. Następnie, następuje ekstrakcja głównych punktów szkieletowych. Wybór cech jest dokonywany w taki sposób, żeby wybrać tylko najbardziej informatywne punkty ciała, co polepsza efektywność numeryczną bez straty jakości. Następnie, klatka referencyjna jest przekształcana biorąc pod uwagę dolny odcinek kręgosłupa. Na koniec, skalowanie typu Min-Max stosuje się dla uzyskania standardowej skali wartości, co umożliwia łatwiejszą integrację z metodami uczenia maszynowego.

Segmentacja akcji oparta na przejściach postur, stosowana zamiast ustalonego okna czasowego, daje tu bardziej naturalną reprezentację ludzkiego ruchu, co jest ważne przy rozpatrywaniu zmienności indywidualnych akcji. Jest to korzystne także dla zapewnienia adaptacyjności systemu.

Przyjęte w pracy rozwiązanie dotyczące ogólnie rozumianego przygotowania danych, głównie przetwarzania wstępnego, zawiera nowoczesne elementy i jest efektywne w przypadku zadania rozpatrywanego w pracy. Jest to bardzo ciekawa część pracy zawierająca oryginalne rozwiązania autorskie Kandydatki.

Duże znaczenie ma tu także właściwe użycie zbiorów danych testowych. Zastosowano głównie zbiór danych z Uniwersytetu Waseda w Tokio, który zawiera dłuższe akcje składające się z kilku ciągle zmieniających się postur, a nie krótkie izolowane akcje, co się często spotyka w innych zbiorach danych testowych dostępnych w sieci.

Treść niniejszego rozdziału i zaproponowane w nim elementy są oryginalne i dobrze udokumentowane. Należy to wysoko ocenić jako cenny i w dużym stopniu oryginalny dorobek Kandydatki.

Rozdział piąty, „Chapter 5: 3D Skeleton-Based Human Posture Recognition Using Unsupervised Learning. Self-Organizing Maps (SOMs)”, jest pierwszym z rozdziałów pracy, które pokazują konkretne rozwiązanie analityczne i algorytmiczne dla jednego z najważniejszych zadań składających się na problem będący przedmiotem pracy,

a mianowicie klasteryzacji ludzkich postur na podstawie trójwymiarowych danych szkieletowych. Celem jest oczywiście identyfikacja i pogrupowanie ludzkich postur na podstawie ich podobieństw strukturalnych, co będzie potem oczywiście miało znaczenia dla zadań związanych z predykcją akcji.

Kandydatka stosuje tu znaną i popularną metodę, tzw. mapę samoorganizującą SOM (Self-organized Map), zwana również samoorganizującą mapą cech SOFM (Self-organized Feature Map), wprowadzoną wiele lat temu przez Teuvo Kohonena. Jest to bardzo dobry wybór dla rozpatrywanego zadania i Kandydatka otrzymała bardzo dobre wyniki. Może jedynie dobrze byłoby bliżej wyjaśnić, dlaczego – po pierwsze – wybór podejścia nienadzorowanego, a – po drugie – metody opartej na SOM jest tu uzasadnionym, czy może jedynym wyborem. Otóż, kandydatka wspomina, że zastosowanie tej metody SOM nie wymaga danych etykietowanych, co może być trudne i kosztowne, i to jest prawda, ale powstaje oczywiście naturalne pytanie, czy np. ten aspekt ekonomiczny jest najważniejszy, czy po prostu etykietowanie danych ludzkich postur jest trudne, zwłaszcza gdy trzeba to robić w czasie rzeczywistym.

Kandydatka używa tu znów dwu zbiorów danych, jak poprzednio, czyli zbioru W-10 i zbioru danych z Uniwersytetu Waseda w Tokio, co jest dobrym rozwiązaniem. Te zbiory danych zapewniają potem dobrą podstawę do analizy różnych aspektów wyników, przede wszystkim właściwej liczby klastrow. Do tego celu Kandydatka zaproponowała ciekawy wskaźnik dyskryminacyjny, tzw. „Discriminant Score (DS)” oparty na separowalności klastrow i ich zwartości. To ciekawa i oryginalna koncepcja dla rozpatrywanego zadania. Aby, porównać wyniki uzyskiwane z użyciem DS, Kandydatka przeprowadziła testy numeryczne na dość dużą skalę, porównując z wynikami uzyskiwanymi przy użyciu innych, często używanych w literaturze wskaźników jakości klasteryzacji, jak np. tzw. Silhouette Coefficient (SC), wskaźnik Dunna (DI), wskaźnik Calińskiego–Harabasa (CH), i wskaźnik Daviesa–Bouldina (DBI). To bardzo ciekawa i wartościowa część rozprawy. Co ważne, Kandydatka użyła do porównania swojego wskaźnika praktycznie wszystkich bardziej popularnych wskaźników klasteryzacji, może jedynie brakuje mi użycia wskaźnika Rand, w kilku wersjach.

Nie ulega wątpliwości, że wyniki uzyskane w tym rozdziale mają duże znaczenie dla osiągnięcia jednego z głównych celów rozprawy, czyli uzyskania dobrych wyników dotyczących predykcji akcji w systemie typu HAP.

Rozdział 6: „Chapter 6: 3D Skeleton-Based Human Posture Classification Using Supervised Learning: Transformers Neural Networks”, jest bardzo dobrym i ciekawym rozdziałem, w którym korzysta się już w pełni z tego, co oferuje współczesna informatyka, a właściwie sztuczna inteligencja. Rozdział ten jest poświęcony klasyfikacji postur ludzkich z użyciem uczenia nadzorowanego, stosując sieć neuronową typu Transformer, nowego rodzaju sieć neuronową do przetwarzania dużych danych sekwencyjnych, która w ostatnich latach cieszy się wielkim zainteresowaniem, zwłaszcza, że stoi ona w dużym stopniu za sukcesami wielu sztandarowych zastosowań, m.in. w tzw. wielkich modelach językowych (LLM- Large Language Models).

Sam pomysł Kandydatki użycia sieci neuronowej typu Transformer jest już bardzo dobry, wychodzący naprzeciw nowym tendencjom, ale dodatkowo wprowadziła ona 3 architektury: Pose Sequence Transformer (PoSeqTNet), Central Posture Transformer (CPT), and Isolated-Attention Transformer (IAT). Jak same nazwy wskazują, dotyczą one podstawowych aspektów związanych z analizą i klasyfikacją postur. Co ważne, Kandydatka zaproponowała

tu kompleksowy system testowania tych architektur, obejmujący m.in. przygotowanie danych i procedury trenowania i testowania. Jest to niewątpliwie cenna część pracy, umożliwiającą kompleksowe przebadanie wyników otrzymanych z użyciem tych 3 architektur. Do testowania Kandydataka użyła przede wszystkim już stosowanych zbiorów danych WUT-10 i z Uniwersytetu Waseda w Tokio, a także dodatkowo zbioru PKU-MMD. Testowanie dotyczyło m.in. dokładności klasyfikacji i odporności na zmieniające się warunki, m.in. na pozycje ciała w sekwencjach akcji, zmieniające się przypadkowo i sekwencyjnie, różnych kątów ustawienia kamer itp.

Wyniki w tej ważnej części pracy są bardzo ciekawe, a przede wszystkim pokazują, że podejścia oparte na sieciach neuronowych typu Transformer (a zwłaszcza IAT – Isolated-Attention Transformer) dają bardzo dobre wyniki przy rozpoznawaniu postur ludzkich.

To bardzo ważny rozdział w pracy, dający niewątpliwie wskazówki dla przyszłych badań.

Rozdział siódmy: „Chapter 7 - An Orientation-Invariant Human Posture Clustering”, jest bardzo ważnym rozdziałem pracy, który już głębiej wchodzi w istotę problemu rozpatrywanego w pracy. W tym rozdziale wprowadza się mianowicie nowe podejście do klasteryzacji postur ludzkich, ale przy założeniu inwariantności od kąta widzenia (np. kamery). Jest to oczywiście niezwykle ważne, bo obserwacja sekwencji postur zwykle zachodzi pod różnymi kątami, z różnych pozycji itp., ponieważ na przykład kamera i jej obiektyw mogą być różnie ustawione, a do tego jeszcze ciała ludzkie mogą być różnej wielkości. Nowa metoda zaproponowana w pracy wprowadza lokalne kąty i względne odległości między punktami ciała, wprowadzane potem do SOM. W pracy pokazuje się, z użyciem wizualizacji, że zaproponowana metoda zapewnia jasną separację klastrów i umożliwia interpretowalne grupowanie postur, dla różnych kątów widzenia. Jasne jest, że w przypadku złożonych, tzw. wielofazowych ruchów postur, napotyka się jednak trudności.

Ogólnie, ważne wyniki otrzymane w tym rozdziale mogą bez wątpienia stanowić punkt wyjścia dla złożonych analiz szkieletowych, inwariantnych od kąta widzenia, co potem będzie mogło być wykorzystane w rozwiązywaniu dalszych zadań, jak np. modelowanie bardziej zaawansowanych ruchów czy rozpoznawanie sekwencji akcji.

Rozdział ósmy, „Chapter 8: Human Action Prediction. SOMs versus Transformer-Based Systems”, to bardzo ważny i ciekawy rozdział, nie tylko zresztą dla tej rozprawy. Otóż, w tym rozdziale rozważa się kompletny system predykcyjny, w którym są moduły mające za zadanie zebranie danych i rozpoznanie postur z danych RGB-D, co opisano w Rozdziałach 4, 5 i 6, uzyskując dobre wyniki. W niniejszym rozdziale, Kandydatka idzie dalej i wprowadza bardziej ustrukturalizowane i interpretowalne podejście, w którym punkt ciężkości jest na rozpoznawaniu postury jako podstawowym bloku dla predykcji akcji. Kandydatka rozpatruje dwie wersje podejścia, które łączą rozpoznawanie postury z predykcją akcji i oba korzystają z relacyjnej bazy danych. Jedno podejście używa sieci neuronowej typu Transformer (opisanego w Rozdziale 6), co pozwala na modelowanie długookresowych zależności przestrzennych. Otrzymane akcje są porównywane z akcjami w bazie danych. A znów, drugie podejście zastępuje sieć neuronową typu Transformer przez samoorganizującą mapę SOM, co jest szczególnie użyteczne dla analizy wielowymiarowych wejść typu trójwymiarowych danych szkieletowych.

Mówiąc krótko, podejście pokazane w tym rozdziale pozwala połączyć zalety obu podejść. Najogólniej, podejście oparte na SOM zapewnia większą odporność i adaptacyjność, natomiast to oparte na sieci neuronowej typu Transformer zapewnia lepszą precyzję i jest bardziej zrozumiałe dla użytkownika. To oczywiście sugeruje, że najlepszym rozwiązaniem byłby system hybrydowy łączący te dwa podejścia i to podejścia Kandydatka potem realizuje.

Podsumowując, jest to też bez wątpienia bardzo wartościowy i cenny rozdział, zwłaszcza w obliczu lawinowego wzrostu zastosowań różnych podejść typu głębokiego uczenia i głębokich sieci neuronowych we wszelkich zadaniach.

Rozdział dziewiąty, „Chapter 9: A Hybrid System of SOM and Transformer Architectures for Human Action Prediction”, jest właśnie bardzo logiczną konsekwencją wyników otrzymanych w poprzednich rozdziałach, czyli zaproponowania dwu różnych podejść do rozpoznawania postur: opartego na SOM i opartego na sieci neuronowej typu Transformer, z których każda ma zalety i wady, ale - na szczęście - są one komplementarne, co w sposób jasny sugeruje, że należy zaproponować zintegrowany system hybrydowy, łączący zalety tych dwu. I tak też się w tym rozdziale dzieje, Kandydatka proponuje architekturę, algorytmikę i implementację takiego systemu hybrydowego, który w sposób zintegrowany łączy w sobie rozpoznawanie postur realizowane przez podejścia oparte na SOM i sieci neuronowej typu Transformer, tak aby dojść do systemu predykcji akcji. Jest to, jak się okazuje, bardzo dobre podejście, które jest oryginalne w tym sensie, że inne rozwiązania hybrydowe łączące w sobie różne podejścia (np. łączące sieci neuronowe konwolucyjne i rekurencyjne) nie dotyczyły rozpoznawania postur. Ciekawym, oryginalnym elementem jest też propozycja właściwej agregacji wyników otrzymanych z tych dwu podejść, a zaproponowane w pracy wskaźniki przekonania (confidence scores) są też bardzo ciekawą propozycją. Są one na razie wprowadzane nieco ad hoc, ale wydaje się, że w przyszłych badaniach można ten element pracy bardzo rozwinąć, opierając się na bogatej literaturze dotyczącej różnego rodzaju współczynników pewności, przekonania, ufności itp., zwłaszcza, że są one często inaczej rozumiane w różnych dziedzinach.

Rozdział dziesiąty, „Chapter 10: Conclusions and Future Works”, jest bardzo dobrze napisanym rozdziałem w ciekawym układzie, a mianowicie podaje się krótki, krytyczny opis głównych cech rozpatrywanych zadań, a potem głównych wyników uzyskanych przez Kandydatkę. Podaje się także problemy i wyzwania stojące przed tego typu badaniami. A w końcu, Kandydatka podaje własną ocenę znaczenia pracy i uzyskanych wyników. Podaje też uwagi na temat dalszych możliwych badań w tematyce rozprawy. Jest to zrobione dobrze, choć - moim zdaniem - Kandydatka jest miejscami zbyt skromna w ocenie swych osiągnięć i oryginalności uzyskanych wyników.

Praca kończy się bogatym wykazem literatury, obejmującym praktycznie wszystkie ważniejsze pozycje związane z tematyką pracy i nowymi podejściami w tym zakresie.

Podsumowując, można stwierdzić, że praca:

- Dotyczy bardzo ważnego z punktu widzenia naukowego i zastosowaniowego problemu zastosowania nowoczesnych narzędzi automatyki, robotyki, informatyki itp. w sensie zarówno analitycznym i algorytmicznym, ale też i implementacji, do analizy i predykcji ludzkich akcji w robotach wspomagających,

- Dotyczy ogólnie problemu robotów wspomagających, co ma wielkie znaczenie społeczno-ekonomiczne w obecnych czasach, gdy we wszystkich praktycznie krajach rozwijających się następują bardzo szybkie starzenie się społeczeństwa,
- Rozpatrywany w pracy problem stanowi przy tym zarówno wyzwanie badawcze, algorytmiczne, jak też implementacyjne i wdrożeniowe,
- Ponadto, problem ten stanowi – co jest powszechne w dzisiejszych czasach – wyjątkowy przykład problemów multidyscyplinarnych, na przecięciu automatyki i robotyki, ale też nauk o danych, algorytmiki, informatyki, itp., co znacznie komplikuje jego rozwiązanie,
- W pracy zaproponowano nowatorskie, oryginalne rozwiązania rozpatrywanego problemu jego części składowych, ale też autorską modyfikację podejść znanych z literatury, zarówno w sensie architektury i koncepcji, jak też algorytmizacji i implementacji.
- Praca jest wyjątkowa w tym sensie, że zawiera zarówno nowe, innowacyjne rozwiązania koncepcyjne i algorytmiczne, jak też nowatorski projekt systemu typu HAP,
- Otrzymane wyniki w pełni potwierdzają zarówno efektywność podejścia Kandydatki, jak ich użyteczność w praktyce,
- Praca jest kompletna w sensie użycia właściwych środków sprzętowych
- i programowych zarówno w fazie projektowania i implementacji, ale też bardzo umiejętnego połączenia wiedzy z wielu dziedzin.

Kandydatka:

- Wykazała się dużą wiedzą w zakresie nowoczesnej literatury przedmiotu i nowych tendencji znanych z literatury, ale też dobrze wybrała tematykę i zakres swojego oryginalnego wkładu badawczego oraz implementacyjnego,
- Posiada niewątpliwie umiejętność samodzielnego prowadzenia pracy naukowej.

Ponadto, należy bez wątpienia podkreślić wielką rolę promotorki, p. prof. dr hab. inż. Teresy Zielińskiej, zarówno poprzez nakierowanie Kandydatki na tak ważną i obiecującą tematykę, jak też w procesie planowania i realizacji tej ambitnej pracy.

W związku z powyższym, stwierdzam, że – moim zdaniem – rozprawa **spełnia wszelkie wymagania ustawowe i zwyczajowo przyjęte w polskim środowisku naukowym**, stawiane rozprawom doktorskim, a także **wymagania stawiane w przesłanym mi pytaniach Rady Dyscypliny**, i wnoszę o jej **przyjęcie oraz dopuszczenie do publicznej obrony**.

Jednocześnie, ze względu na wspomniane już zalety pracy, proponowałbym, żeby w przypadku odpowiedniego przebiegu obrony i innych etapów postępowania, a także spełnienia odpowiednich wymagań przyjętych w Politechnice Warszawskiej, rozpatrzyć jej **wyróżnienie**.

J. Kępczyński