

Recenzja pracy doktorskiej

Autorka pracy doktorskiej:	mgr inż.. Ekemeyong Awong Lisiane Esther
Tytuł:	Predicting Human Actions for Utilisation in Assistive Robots
Przewód doktorski w:	Politechnika Warszawska, wydział MEiL
Dyscyplina naukowa:	Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne
Promotor:	Prof. dr hab. inż. Teresa Zielińska
Promotor pomocniczy:	Dr. Vibekananda Dutta
Recenzent:	dr hab. Inż. Piotr Szynkarczyk, Instytut Łuksiewicz - PIAP

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 podstawie tej Uchwały zostałem wyznaczony jako recenzent rozprawy doktorskiej mgr inż. Ekemeyong Awong Lisiane Esther pt. „Predicting human actions for utilisation in assistive robots”

1. Ogólne streszczenie rozprawy

Recenzowana niniejszym rozprawa doktorska przedstawia opis programu badawczego, dotyczącego przewidywania działań człowieka (HAP), ze szczególnym uwzględnieniem zastosowania w aplikacjach robotycznych, które muszą działać w bliskiej fizycznej i społecznej bliskości z ludźmi. Praca pozycjonuje się na styku automatyki, komputerowej analizy obrazu, uczenia maszynowego i interakcji człowiek-robot, a jej głównym celem jest umożliwienie programom sterującym robotami przewidywania działań człowieka odpowiednio wcześniej i wiarygodnie. Główna teza kandydatki przedstawiona jest w rozdziale 3. Celem tych badań jest opracowanie wydajnego, adaptowalnego i opartego na danych systemu HAP, który wykorzystuje rozpoznawanie postawy człowieka – zarówno z wykorzystaniem metod uczenia nadzorowanego, jak i nienadzorowanego – do przewidywania działań człowieka w czasie rzeczywistym. Celem jest także przezwyciężenie ograniczeń obecnych systemów poprzez poprawę generalizacji, zmniejszenie zależności od dużych zbiorów danych oraz umożliwienie przewidywania działań człowieka na (relatywnie) dłuższe perspektywy czasowe.

Rozprawa składa się z dziesięciu rozdziałów, które obejmują zagadnienia od motywacji i tła (rozdziały 1–2), przez akwizycję, wstępną obróbkę i filtrowanie danych (rozdział 4), po oparte na uczeniu rozpoznawanie postaw (HAP) za pomocą nienadzorowanych map samoorganizujących (SOM) i nadzorowanych architektur transformatorów (rozdziały 5 i 6), deskryptory postawy człowieka niezależne względem punktu widzenia (rozdział 7), a na końcu coraz to bardziej rozbudowane systemy HAP (rozdziały 8 i 9). Rozdział 10 zawiera wnioski i propozycję zakresu dalszych badań.

Wstęp (rozdział 1) podaje przykłady przewidywanych praktycznych zastosowań opisywanej w pracy metodyki przewidywania działań człowieka poprzez przekonujące przykłady z: opieki zdrowotnej, nadzoru bezpieczeństwa, systemów jazdy autonomicznej, sportu i rozrywki, a także innych zastosowań wymagających współpracy ludzi z robotami.

Rozdział 2 zawiera obszerny i starannie opracowany przegląd najnowocześniejszych rozwiązań dotyczących współpracy ludzi z robotami. Sekcja 2.3 porównuje metody wykrywania – noszone czujniki IMU, RFID, Wi-Fi, wizyjne i RGB-D. Sekcja 2.4 analizuje zestawy gotowych baz danych zachowań ludzkich. Sekcja 2.5 porządkuje podejścia analityczne z użyciem m.in. sztucznych sieci neuronowych, LLM (Large Language Models), SVM (Support Vector Machines), modele probabilistycznych i drzew decyzyjnych. Sekcja 2.6 analizuje otwarte wyzwania, takie jak zależność od punktu obserwacji czy niedobór danych do wczesnego przewidywania. Wskazana jest potrzeba budowania modeli o wysokiej wydajności obliczeniowej.

Rozdział 3 formułuje cel badań i tezy robocze.

Rozdział 4 to kompleksowe omówienie metod akwizycji i wstępnego przetwarzania danych. Sekcja 4.2.1 szczegółowo opisuje kamery głębi ZED 2 i Microsoft Azure Kinect, a tabela 4.1 porównuje ich właściwości istotne dla estymacji konfiguracji tzw. szkieletu (postaci ludzkiej). Rysunki 4.2–4.6 ilustrują surowe tryby RGB, głębokości i szkieletu, segmentację sekwencji akcji oraz proces ekstrakcji cech, transformacji układu odniesienia i normalizacji (sekcje 4.3.1–4.3.6).

Rozdział 5 wprowadza modele rozpoznawania postawy (SOM: Self-Organizing Maps) do klasteryzacji postawy. Podstawą metodologiczną rozdziału jest tzw. wynik (wskaźnik) dyskryminacyjny (Discriminant Score, DS), zaproponowany jako uzupełnienie klasycznych wskaźników trafności

klastryzacji. Rysunki 5.1–5.4 porównują wskaźniki Silhouette, Dunna, Daviesa–Bouldina i Calinskiego–Harabasz dla różnych liczebności klastrów kandydujących. Rysunki 5.5–5.13 analizują zachowanie wskaźnika DS w kontekście metryk cosinusowych i euklidesowych, a rysunki 5.14–5.18 śledzą dynamikę treningu, rozkłady próbek w klastrach oraz lukę między aktywacjami BMU (Best Matching Unit of SOM) i drugiej BMU. Analiza rozpoczyna się od kontrolowanych podzbiorów (sekcja 5.8) w celu kalibracji podejścia, a następnie przechodzi do pełnego zbioru danych (sekcje 5.10–5.12), gdzie wskaźnik DS pomaga ustabilizować wybór „kluczowych postaw” postaci ludzkiej, które najlepiej dzielą przestrzeń działania.

Rozdział 6 omawia uczenie nadzorowane z architekturą Transformer zoptymalizowaną pod kątem analizowania sekwencji postaw szkieletu 3D (postaci ludzkiej). Badane są trzy architektury: PoSeqTNet (rys. 6.2), Central Posture Transformer (CPT; rys. 6.7) oraz Isolated-Attention Transformer (IAT; rys. 6.8). Program badań obejmuje sekwencyjne i losowe porządkowanie danych wejściowych, różne długości sekwencji i zestawy danych, metryki dla poszczególnych klas oraz wzorce zachowań obserwowanego człowieka. Tabela 6.4 zestawia architektury (modele) z danymi z okresu uczenia, które później wpływają na decyzje dyskutowane w rozdziale 8. Rozdział ten przedstawia przekonujący argument za używaniem modelowania postawy opartym na uwadze.

Rozdział 7 opisuje metodologię ekstrakcji cech niezmienniczych względem orientacji dla klasteryzacji postawy człowieka opartej na SOM. Rozdział ten omawia problem zależności od punktu widzenia w rozpoznawaniu postawy i analizie ruchu człowieka, proponując metodę klasteryzacji postawy niezmienniczą względem orientacji. Autorka proponuje metodę, która ekstrahuje cechy niezmiennicze względem widoku z danych szkieletu 3D, w szczególności są to względne odległości między segmentami ciała oraz kąty między segmentami w poszczególnych punktach ciała, które pozostają spójne dla różnych kątów kamery, rozmiarów ciała i zmienności anatomicznej różnych ludzi. Cechy te są wykorzystywane do trenowania sieci SOM opisanej w rozdziale 5. Poprzez serię eksperymentów z wykorzystaniem zbiorów danych WUT-10 i Waseda, autorka wykazuje, że to podejście znacząco poprawiające jakość i spójność klasteryzacji postawy ludzkiego ciała.

Rozdziały 8 i 9 opisują realizację kompletnych systemów HAP z wykorzystaniem opisanych we wcześniejszych rozdziałach metodyk. Rozdział 8 przedstawia proces wykonywania procesów HAP (rysunek 8.1), oraz pokazuje predykcję opartą na SOM i Transformerze, a także interfejsy użytkownika przedstawione na rysunkach 8.3–8.4. Rozdział 9 przedstawia system hybrydowy (rysunek 9.1) z adaptacyjnym ważeniem ufności (rysunek 9.2) i interfejsem, który ujawnia, który model aktualnie zarządza predykcją (rysunek 9.3). Sekcja 9.5.1 opisuje demonstrację robotyczną, w której system przewiduje zbliżającą się czynność pobrania i kieruje platformą mobilną (robot kroczący Spot) do pobrania i dostarczenia obiektu (rysunek 9.5), zamykając w ten sposób w praktyce pełną drogę od percepcji do działania.

Rozdział 10 syntetyzuje wyniki pracy badawczej (sekcje 10.1–10.7) i formułuje przyszłe kierunki badań (sekcja 10.8).

Lista publikacji jest obszerna i dobrze dobrana. Między innymi są to publikacje w „Pomiary Automatyka Robotyka” (2023), analiza klastrowania w „Sensors” (2023), artykuły konferencyjne IFToMM i ICMERR (2023), artykuł w RoMoCo (2024) oraz dalsze manuskrypty będące w trakcie recenzji i druku.

Podsumowując, rozprawa stanowi dogłębną analizę HAP w warunkach realistycznych ograniczeń sensorycznych i obliczeniowych, charakteryzującą się nowatorskim podejściem metodologicznym, rygorystyczną oceną własnych wyników i przekonującymi demonstracjami na poziomie praktycznym.

2. Ocena wiedzy teoretycznej

Wiedza teoretyczna kandydatki robi bardzo dobre wrażenie, jest dogłębna, wnikliwa i wynika z bardzo dobrego opanowania wybranej do pracy doktorskiej tematyki. Rozdział 2 to nie tylko przegląd „state of the art”, jego struktura ma także na celu wyjaśnienie jak ograniczenia sensoryczne przekładają się na wybory modelowania w zakresie przewidywania pozycji i przyszłych działań człowieka.

Prezentacja uczenia się bez nadzoru (rozdział 5) precyzyjnie przedstawia regułę uczenia się SOM, ewoluującą funkcję sąsiedztwa oraz rolę jednostek najlepszego dopasowania (BMU) w mapowaniu przestrzeni cech. Kandydatka omawia tryby awaryjne – takie jak fragmentacja klastra, wrażliwość na inicjalizację i tendencję globalnych indeksów ważności do rozbieżności – i uzasadnia propozycję wyniku (wskaźnika) dyskryminacyjnego (DS) jako lokalną, zorientowaną na decyzje alternatywę rozwiązującą powyższe problemy, która koncentruje się na separowalności wokół BMU – co pokazano w rozdziałach 8 i 9.

Elementy uczenia się z nadzorem (rozdział 6) są zakorzenione we współczesnej teorii modelowania sekwencji. Kandydatka w oparciu o swoją znajomość tematyki wyjaśnia dlaczego kodowanie pozycyjne ma znaczenie w sekwencjach postaw; oraz w jaki sposób wybory architektoniczne systemu analizy HAP wpływają na stabilność treningu i czas uczenia systemu HAP (tabela 6.4). Eksperymenty dotyczące różnic między sekwencyjną a losową kolejnością klatek (rysunki 6.3–6.6; tabele 6.1–6.3) dostarczają empirycznego wsparcia dla teoretycznego potwierdzenia tezy omawianej pracy doktorskiej.

Rozdział 7 demonstrowa solidne zrozumienie omawianych w pracy zagadnień. Przechodząc od absolutnych położań stawów do względnych odległości między segmentami i kątów wyśrodkowanych na stawach (rysunek 7.1), kandydatka dostosowuje swój projekt do niezmienników biomechanicznych. Wynikają z tego udoskonalenia w zakresie opanowania przez system problemów wynikających z różnych punktów widzenia.

Rozdziały dotyczące propozycji systemu HAP (rozdziały 8 i 9) opierają się na wynikach analiz i opisach, a także własnych wnioskach z wcześniejszych rozdziałów. Zaznaczyć należy że prace opisane we wcześniejszych rozdziałach zostały wykonane wyjątkowo starannie i obszernie, stanowiąc bardzo solidny fundament do rozdziałów 8 i 9. Architektury HAP zaproponowane w pracy wynikają z przeprowadzonej dogłębnie pracy przeglądowej, analitycznej i badawczej. Są oparte na wiedzy zgromadzonej w bardzo systematyczny sposób.

Podsumowując - rozprawa spełnia, a nawet przewyższa standard opanowania teorii oczekiwany na poziomie doktorskim. Kandydatka nie tylko zna literaturę, ale także wykorzystuje dogłębnie opanowaną teorię do podejmowania decyzji projektowych i interpretacji wyników empirycznych.

3. Ocena niezależności badań

Dowody na zdolność do prowadzenia niezależnych badań są liczne i przekonujące.

Po pierwsze, kandydatka zbudowała pełen system danych i ewaluacji, zamiast polegać wyłącznie na publicznych zbiorach danych. Rozdział 4 dokumentuje protokoły rejestrowania, kalibrację wieloczułnikową oraz przepływy pracy adnotacji i segmentacji, demonstrując tym samym zbudowanie bazy do prowadzenia własnych eksperymentów. Wybór pracy zarówno z czujnikami ZED 2, jak i Azure Kinect (sekcja 4.2.1) wymagał integracji technicznej w ramach oprogramowania i starannego benchmarkingu jakości otrzymywanej pozycji szkieletu ludzkiej postawy.

Po drugie, same opisane w pracy algorytmy stanowią oryginalny projekt. Wynik (wskaźnik) Dyskryminacyjny (DS) został zaproponowany, sformalizowany i przetestowany pod kątem standardowych indeksów w wielu systemach klastrowania. Podobnie, architektury Transformer zostały sprawnie zaadaptowane do systemu HAP.

Po trzecie, gotowość kandydatki do wystawiania swoich modeli na niekorzystne warunki – losowe kolejności ramek, zmienne długości sekwencji i generalizację między zbiorami danych – świadczy o niezależnym, krytycznym nastawieniu.

Po czwarte, etapy integracji systemu (rozdziały 8–9) wykazują inicjatywę wykraczającą daleko poza prototypowanie algorytmów. Co szczególnie istotne, demonstracja robotyczna (sekcja 9.5.1; rysunek 9.5) dowodzi zdolności kandydatki do przekształcania modeli laboratoryjnych w realne działające systemy – niezbędny krok w robotyce, który obejmuje ograniczenia sprzętowe, względy bezpieczeństwa oraz wszelkie inne problemy wynikające z przekształcenia teorii w praktykę.

Po piąte, lista publikacji wymieniona na początku rozprawy potwierdza stałe zaangażowanie w dialog ze społecznością naukową w czasopismach (Sensors; Pomiar Automatyka Robotyka) i na konferencjach międzynarodowych (IFTToMM; ICMERR; RoMoCo). Ta zewnętrzna walidacja zazwyczaj wymaga od autora obrony decyzji metodologicznych, reagowania na krytykę ze strony innych naukowców i odpowiedniego korygowania – działań, które kształtują niezależność osądu.

Na koniec, sekcje poświęcone dyskusji i ograniczeniom (rozdział 10) prezentują odpowiedni poziom autorefleksji badawczej, a nie jedynie oświadczenie o osiągnięciu sukcesu. Kandydatka wskazuje i uznaje ograniczenia (np. scenariusze jednoosobowe w niektórych eksperymentach, wysiłek związany z potrzebą adaptacji bazy danych) i proponuje konkretne rozwiązania (m.in. bogatsze zbiory danych, w tym wieloosobowych; ściślejsze powiązanie detekcji i śledzenia). Podsumowując, rozprawa wykazuje autonomię w zakresie wyboru problemu, rozwoju metodologicznego, projektowania eksperymentów i upowszechniania wyników.

4. Ocena oryginalności

Rozprawa zawiera wiele warstw oryginalności, które razem stanowią znaczący wkład w dyscyplinę. Na poziomie metrycznym, Wynik (wskaźnik) Dyskryminacyjny (Discriminant Score – DS) opisany w rozdziale 5 stanowi trafną odpowiedź na problem praktyczny: niestabilność i sporadycznie sprzeczne wskazówki oferowane przez globalne wskaźniki trafności klastrowania w odniesieniu do przestrzeni pozycji postawy ludzkiego ciała. DS zapewnia narzędzie do wyboru liczby klastrowania i diagnozowania niejednoznacznych postaw. Analizy porównawcze w odległościach cosinusowych i euklidesowych (rysunki 5.5–5.13) oraz walidacja zarówno na podzbiorach kontrolowanych, jak i pełnych zbiorach danych (sekcje 5.8–5.12) potwierdzają tezę o nowości i użyteczności pracy.

Na poziomie architektonicznym, trio modeli Transformer – PoSeqTNet, CPT i IAT – jest trafnie dostosowane do dziedziny rozpoznawania postawy szkieletu. PoSeqTNet kładzie nacisk na uporządkowanie czasowe; CPT wyróżnia centralną postawę jako punkt zaczepienia, aby zmniejszyć zmienność sekwencji; IAT bada efekt izolowania uwagi w podzbiorach. Wyniki związanych z tym analiz i eksperymentów wnoszą nową wiedzę na temat tego, jak należy konfigurować mechanizmy uwagi w celu rozpoznawania postawy i jej przewidywania.

Na poziomie reprezentacji, cechy niezmiennie względem orientacji z rozdziału 7 stanowią przemyślaną syntezę intuicji biomechanicznej i pragmatyzmu obliczeniowego. Premiując odległości względne i kąty centrowane na stawach nad współrzędnymi bezwzględными, kandydatka pokazuje, że można osiągnąć znaczne korzyści w generalizacji widoku obserwowanego człowieka bez uciekania się do skomplikowanych procesów dopasowania 3D. Jest to oryginalny wkład o wyraźnej wartości praktycznej.

Oryginalność pracy potwierdzają także recenzowane publikacje obejmujące artykuły dotyczące badań, metod i systemów, z wyraźnym wkładem autorskim kandydatki.

Połączenie innowacji metrycznych, dogłębnej analizy problemu oraz implementacja praktyczna i jej badania eksperymentalne stanowią spójny i oryginalny wkład w HAP i robotykę wspomagającą.

5. Ocena końcowa i rekomendacja

Mocne strony

Rozprawa wyróżnia się;

- (1) jasnym sformułowaniem problemu opartym na rzeczywistych scenariuszach asystujących;
- (2) opartą na zasadach metodologią, która równoważy interpretowalność i dokładność; (
- (3) dbałością o jakość danych i ich przetwarzanie;
- (4) rygorystycznymi, dobrze udokumentowanymi eksperymentami i ich weryfikacją
- (5) integracją na poziomie praktycznie wykonanego systemu, z demonstracją robotyczną.

Tekst jest uporządkowany i precyzyjny, a rysunki/tabele są liczne i czytelne (np. rysunki 4.2–4.6; 5.1–5.18; 6.2–6.11; 7.1–7.3; 8.1–8.4; 9.1–9.5; tabele 4.1; 6.1–6.9; 8.1–8.2).

Praca odzwierciedla również dobrą metodykę i strukturę badań: opisano zbiory danych, sformułowano założenia i wskazano ograniczenia.

Ograniczenia i konstruktywne sugestie

Po pierwsze, zależność od dokładnej ekstrakcji szkieletu pozostaje wąskim gardłem, szczególnie w scenach zatłoczonych lub zasłoniętych.

Po drugie, chociaż warstwa wnioskowania oparta na bazie danych zapewnia interpretowalność, to skalowanie do bardzo dużych katalogów działań może wymagać reprezentacji opartych na ontologii (por. rysunek 9.4) i ustrukturyzowanych a priori, jak zresztą sugeruje sama kandydatka w rozdziale 10.

Po trzecie, w przypadku wdrażania w systemach pomocy krytycznej dla bezpieczeństwa, szacowanie niepewności i zachowania zabezpieczające przed awariami powinny być bardziej precyzyjnie modelowane (np. skalibrowane wskaźniki zaufania, polityka wstrzymywania się od głosu lub współprzekazywanie kontroli w przypadku spadku zaufania).

Sugestie te nie umniejszają obecnych osiągnięć; nakreślają one ścieżkę badań podoktoranckich.

Wpływ i znaczenie

Wkład omawianej pracy ma wyraźne znaczenie dla praktycznych zastosowań robotyki, takich jak opieka zdrowotna, produkcji w zespołach człowiek-robot i wielu innych.

Praca ta potencjalnie może mieć wpływ na stan wiedzy w kontekście poszukiwania wiarygodnych systemów sztucznej inteligencji (autonomii) dla sterowania ucieleśnionych agentów. Zaproponowany system hybrydowy stanowi konkretny przykład rozwiązywania złożonych problemów związanych z działaniem robotów w świecie rzeczywistym i przy interakcji z ludźmi.

Formalny wniosek

Przedstawiona do recenzji rozprawa doktorska stanowi oryginalne rozwiązanie istotnego problemu naukowego w dyscyplinie naukowej Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne oraz potwierdza wiedzę teoretyczną i umiejętność samodzielnego rozwiązywania problemów naukowych przez Doktorantkę. Rozprawa w pełni spełnia ustawowe wymagania dotyczące rozpraw doktorskich określone w artykule 187 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce.

Jednocześnie, ze względu na wyjątkowo solidną strukturę pracy, bardzo dobre opisanie state-of-the-art, zastosowaną imponująco spójną i dogłębną metodykę badawczą, doprowadzenie rozważań teoretycznych do badań rzeczywistego systemu robotycznego, a także znaczenie wyników dla rozwoju kwestii interakcji człowiek-robot, proponuję, aby w przypadku odpowiedniego przebiegu obrony i innych etapów postępowania, a także spełnienia odpowiednich wymagań przyjętych w Politechnice Warszawskiej, rozpatrzyć jej wyróżnienie.


25.8.2025

