

Tytuł, stopień, imię i nazwisko  
prof. dr hab. inż. Bogdan Kwolek



data 3.09.2020 r.

PRZEWODNICZĄCY RADY DISCYPLINY  
Automatyka, Elektronika i Elektrotechnika

*Szeląg*  
prof. dr hab. inż. Wojciech Szelaq

### RECENZJA ROZPRAWY DOKTORSKIEJ

przygotowana dla Rady dyscypliny automatyka, elektronika i elektrotechnika  
Politechniki Poznańskiej

**Tytuł rozprawy: A new probabilistic approach to global localization in robotics**

**Autor rozprawy: mgr inż. Jan Wietrzykowski**

**Promotor: prof. dr hab. inż. Piotr Skrzypczyński**

1. Przedstawiona mi do recenzji rozprawa doktorska dotyczy globalnej lokalizacji agenta wewnątrz pomieszczeń. Prace badawcze obejmowały projektowanie oraz budowę kompletnych systemów dla inteligentnych robotów. Globalna lokalizacja agenta realizowana jest w oparciu o segmenty płaszczyzn. Zasadnicza część badań naukowych ukierunkowana była na projektowanie algorytmów, które zapewniają rozpoznanie pozycji bez błędnych rozpoznań, tzn. fałszywie dodatnich rozpoznań. W pracy zaprojektowano oraz eksperymentalnie przebadano dwa systemy: system operujący na danych z kamery RGB-D oraz system operujący na obrazach z kamer w układzie stereoskopowym. Proponowane algorytmy budują funkcję gęstości prawdopodobieństwa (PDF), która reprezentuje przekonanie agenta o jego pozycie. Wspomniana funkcja budowana jest w oparciu o lokalne/częściowe przesłanki, które obarczone są niepewnością. Przesłanki wynikają z wykrywanych segmentów płaszczyzn. W rozprawie zaproponowano nowatorskie i zarazem konkurencyjne rozwiązania do budowy i zarządzania zarówno mapą globalną jak i mapami lokalnymi. Dzięki temu, że kamery RGB-D umożliwiają precyzyjną rekonstrukcję sceny 3D, w systemie opartym o takie kamery wyznaczanie pozycji odbywa się z wykorzystaniem równań nieskończonych płaszczyzn. W omawianym rozwiązaniu kandydaci do dopasowania wyszukiwani są w oparciu o deskryptory oparte o histogramy kolorów. Celem wykorzystania potencjału kamer w układzie stereoskopowym w zadaniach globalnej lokalizacji agenta zaproponowano metodę do wykrywania segmentów płaszczyzn. Wykrywanie segmentów płaszczyzn realizowane jest w oparciu o autorską architekturę głębokiej sieci neuronowej. Dzięki zaproponowanej niezależnej od parametrów kamery reprezentacji wektorów normalnych uzyskano poprawę dokładności rekonstrukcji geometrii. Pobieranie kandydatów do dopasowania z globalnej mapy realizowane jest w oparciu o generowane deskryptory segmentów płaszczyzn przez dedykowaną gałąź w omawianej architekturze sieci neuronowej. Estymacja pozycji odbywa się z wykorzystaniem zaproponowanej niepewności estymacji głębi, a także autorskich granic segmentów płaszczyzn. Wynikiem końcowym rozprawy jest kompletny system, który jak pokazano empirycznie, na wymagającym repozytorium obrazów na potrzeby globalnej lokalizacji agenta wewnątrz pomieszczeń, przy koniecznym warunku uzyskania bezbłędnych rozpoznań, uzyskał lepsze wskaźniki rozpoznawania pozycji. Wyniki badań empirycznych pokazują, że zaproponowana sieć umożliwia wykrywanie segmentów z większą skutecznością, zaś dokładność rekonstrukcji sceny jest wyższa od dokładności uzyskiwanej przez konkurencyjne algorytmy. Dzięki zaproponowanym deskryptorom

wyznaczanie pozy odbywa się w oparciu o mniejszą liczbę kandydatów do dopasowania w porównaniu do rozwiązań opartych o histogramy kolorów. Praca ma charakter badawczo-doświadczalny. Autor zaproponował oryginalne rozwiązania oraz zbudował kompletne systemy. W ramach prac eksperymentalnych zgromadzono repozytoria posegmentowanych obrazów, które wraz z opracowanym oprogramowaniem udostępniane są w sieci Internet. Zasadniczym celem naukowym było opracowanie konkurencyjnych rozwiązań na potrzeby globalnej lokalizacji agenta wewnątrz pomieszczeń. Teza pracy wyraża się w przekonaniu iż lokalne, częściowe i niepewne cechy reprezentujące segmenty planarne pozwalają zbudować funkcję gęstości prawdopodobieństwa, opisującą globalną metryczną pozę agenta w środowisku stworzonym przez człowieka. Teza pracy jest konsekwentnie rozwijana w rozprawie. Jej prawdziwość została wykazana empirycznie. Tematyka rozprawy jest w pełni uzasadniona, interesująca i aktualna. Wyniki badań eksperymentalnych są przekonujące.

2. Rozprawa doktorska Pana mgr. inż. Jan Wietrzykowskiego przygotowana została jako cykl sześciu prac naukowych. Trzy prace składające się na cykl zostały opublikowane w renomowanych czasopismach z listy JCR o wysokich współczynnikach wpływu (IF=3.5, 3.7, 3.0), dużej liczbie punktów ministerialnych (140, 200, 100), a także na jednej uznanej konferencji międzynarodowej, która jest bardzo wysoko punktowana na liście ministerialnej (140 pkt.). Łącznie opublikowano sześć artykułów w znaczących czasopismach naukowych (J. of Aut. Mobile Robotics and Intell. Syst., 2016, Rob. and Autonomous Syst., 2019, IEEE Rob. and Aut. Letters, 2022, IEEE Access, 2022) oraz zaprezentowano na znaczących konferencjach (European Conf. on Mob. Rob., 2017, IROS, 2021). Dwie spośród opublikowanych prac są pracami autorskimi, zaś w czterech pozostałych publikacjach składających się na cykl, będących podstawą rozprawy, Doktorant jest pierwszym autorem. Prace realizowane były ze wsparciem grantu NCN Preludium oraz grantu Unii Europejskiej TERRINet (EU H2020). Zakres prac badawczych świadczy o bardzo dużym zaangażowaniu i samodzielności Autora. Jednocześnie skład autorski tych publikacji wskazuje na duże wsparcie i współdziałanie zespołu badawczego w realizacji badań naukowych, których wyniki zawarte są w niniejszej rozprawie doktorskiej. Publikacje składające się na cykl są na bardzo wysokim poziomie. Wysoki poziom artykułów wynika nie tylko z bardzo dobrej prezentacji wyników, a w szczególności z uzyskania konkurencyjnych rozwiązań i wyników, ale także z tego, że proponowane rozwiązania mają podstawy teoretyczne. Podobnie wysoką jakością jak publikacje odznacza się streszczenie oraz podsumowanie rozprawy. Autor w bardzo przystępnej, syntetycznej formie przedstawił całość zagadnienia, naświetlił cele i przebieg badań oraz podkreślił wagę uzyskanych wyników, wskazując na wkład w badania podstawowe jak i w rozwój systemów robotyki o otwartym kodzie źródłowym, a także w ważne zagadnienie ewaluacji systemów robotycznych. W bibliografii Autor przywołał 33 prace, w tym 6 prac wchodzących w cykl publikacji. Układ rozprawy doktorskiej Pana mgr. inż. Jana Wietrzykowskiego jest zgodny z normami przyjętymi dla tego typu opracowań. Jest on podporządkowany przyjętej tezie badawczej oraz weryfikacji empirycznej hipotez badawczych. Rozprawa doktorska świadczy o dojrzałości badawczej Doktoranta, ukazuje bogaty warsztat badawczy, umiejętność interpretacji wyników, ich syntetycznego przedstawienia, a w szczególności umiejętność samodzielnego prowadzenia pracy naukowej. Na podkreślenie zasługuje syntetyczność, systematyczność i przejrzystość wynikająca z właściwie przyjętej koncepcji przygotowania rozprawy.

3. Rozprawa napisana jest w języku angielskim. W początkowej części przedstawiono motywację podjęcia tematu, omówiono zasadnicze wyzwania, postawiono tezę badawczą oraz omówiono wkład własny, a następnie dość szczegółowo omówiono zawartość artykułów naukowych, które opublikowane zostały w trakcie realizacji pracy doktorskiej. W dalszej części rozprawy zamieszczone są prace składające się na cykl publikacji. W końcowej części pracy zawarto wnioski oraz dyskusję. Doktorant dość szczegółowo omówił nowatorskie aspekty w odniesieniu do reprezentacji segmentów płaszczyzn oraz detekcji segmentów płaszczyzn, budowy mapy oraz zarządzania mapą, deskryptora wyglądu, a także wyznaczania pozy. Wskazał także na potencjał przygotowanych repozytoriów danych na potrzeby globalnej lokalizacji agenta wewnątrz pomieszczeń. Na końcu prezentowane jest znaczenie rozprawy oraz dalsze prace.

4. Tematyka badań podjęta przez Doktoranta wyniknęła ze szczegółowej analizy ograniczeń istniejących systemów do globalnej lokalizacji agenta wewnątrz pomieszczeń, a także wyzwań, którym to powinny sprostać roboty autonomiczne. Po poprawnym zidentyfikowaniu ograniczeń systemów do globalnej lokalizacji robota, Doktorant rozpoczął prace nad systemami opartymi o płaszczyznę, a następnie segmenty płaszczyzn znajdujące się w pomieszczeniach zamkniętych. Warto wspomnieć, że wówczas w nielicznych pracach zaczęły pojawiać się koncepcje użycia innych cech i deskryptorów niż punktowe. Co więcej, Doktorant zaproponował sformułowanie problemu lokalizacji agenta jako problemu budowy mapy gęstości prawdopodobieństwa, które wyraża jego przekonanie o położeniu, którego to maksimum reprezentuje jego pozę i określane jest względem globalnej mapy płaszczyzn.

W autorskiej pracy opublikowanej w 2016 r. przebadano dwie reprezentacje płaszczyzn na potrzeby globalnej lokalizacji agenta, tj. opartej na parametryzacji SE(3) oraz na parametryzacji analogicznej do kwaternionów, a następnie jednoczesnej optymalizacji pozy i położenia cech w oparciu o solver  $g^2o$  przeprowadzono szereg badań symulacyjnych. Zasadniczym ograniczeniem tego podejścia było ograniczenie zidentyfikowane także w innych pokrewnych pracach, które wynika wprost z natury przyjętych reprezentacji dla translacji i rotacji, w sytuacji gdy pomiary są obciążone szumami. Praca jest wartościowa z uwagi na zaproponowanie i praktyczne przebadanie nowych reprezentacji na potrzeby globalnej lokalizacji robota.

We współautorskiej pracy opublikowanej w materiałach konferencyjnych European Conference on Mobile Robotics zaproponowano nowatorskie podejście do problemu lokalizacji agenta w oparciu o mapę gęstości prawdopodobieństwa (PDF), które wyraża jego przekonanie o położeniu, którego to maksimum wyraża jego pozę i określane jest względem globalnej mapy płaszczyzn. Wspomniana PDF budowana jest w oparciu o prymitywy geometryczne typu płaszczyzny w przestrzeni 6D reprezentującej pozę agenta. Co więcej, do budowy map zaproponowano użycie segmentów płaszczyzn. Budowanie chmur punktów reprezentujących widok sceny lokalnej (mapa lokalna) oraz mapę globalną odbywa się w oparciu o algorytm znany pod nazwą ElasticFusion. Chmury punktów są następnie segmentowane poprzez grupowanie superwokseli, a powstałe w ten sposób płaskie płyty są łączone w większe segmenty. Nowość podejścia polega na wykorzystaniu wielu trójek potencjalnie korespondujących płaszczyzn do wygenerowania funkcji PDF. Wspomniane trójki składają się z trzech par powiązanych cech typu płaszczyzna. Każda trójka podlega ocenie jeśli indukuje potencjalną transformację płaszczyzn w układzie współrzędnych bieżącego widoku czujnika na układ współrzędnych mapy globalnej. Ostateczna hipoteza wyznaczana jest z użyciem wag promujących lepsze dopasowania. Ewaluacja skuteczności rozpoznawania miejsca przeprowadzona została na

wymagającym repozytorium danych. Dane odniesienia (ang. *ground-truth*, GT) zarejestrowano przy pomocy systemu motion capture OptiTrack. Wyniki badań eksperymentalnych wskazują, że po odpowiednim doborze parametrów na wspomnianym zbiorze danych możliwe jest rozpoznanie pozycji bez błędnych rozpoznań. Zbiór danych oraz kody źródłowe systemu są udostępnione w sieci Internet. Omawiana praca jest bardzo wartościowa. Doktorant jest autorem zasadniczych koncepcji, a także opracował oraz przygotował zasadniczą część artykułu.

Powyższa praca została rozszerzona o nowatorskie elementy, a następnie opublikowana w uznanym czasopiśmie *Rob. & Aut. Syst.* Celem zmniejszenia trudności praktycznych wynikających z natury translacji i rotacji zaproponowano nową miarę dopasowania segmentów płaszczyzn. W zmodyfikowanej implementacji systemu mapy lokalne i mapa globalna budowane są bez odwołań do ElasticFusion. Dzięki temu uproszczono integrację modułu globalnej lokalizacji agenta z systemami SLAM. Wyniki uzyskiwane przez system odniesiono do wyników uzyskiwanych przez system ORB-SLAM2. W pracy pokazano m.in., iż aczkolwiek wspomniany system może osiągać nieco lepsze wyniki na fragmentach trajektorii z których widoczne są obrazy sceny bogate w cechy, to autorski system umożliwi lokalizację w oparciu o jedynie ubogie w cechy ściany sceny. W pracach badawczych Autor poprawnie zidentyfikował szereg problemów badawczych, które były przedmiotem dociekań w kolejnych pracach. Doktorant był autorem zasadniczych koncepcji, a także opracował oraz przygotował zasadniczą część artykułu.

W pracy opublikowanej w wysoko punktowanym czasopiśmie *IEEE Rob. & Aut. Lett.* zaproponowano architekturę sieci neuronowej do detekcji segmentów płaszczyzn oraz wyznaczania geometrii na podstawie pary obrazów z kamery stereoskopowej. Architekturę sieci neuronowej oparto na architekturze sieci Plane R-CNN. Geometria między płaszczyznami wyznaczana jest w oparciu o dedykowaną gałąź oraz rozwiązania znane pod nazwą *cost volume*. Propozycja użycia wspomnianej techniki do wyznaczania normalnych i integracji w proponowanej architekturze sieci jest oryginalna. Uczenie sieci odbywało się na zbiorze syntetycznym wygenerowanym w niniejszej pracy, a także na zbiorze zarejestrowanym w projekcie TERRINet. Opracowana i wytrenowana sieć Stereo Plane R-CNN zapewnia nie tylko znacznie lepszą jakość w zakresie estymacji wektorów normalnych, a także dzięki możliwości użycia kamer w układzie stereoskopowym stwarza nowe możliwości, które nie są dostępne dla kamer RGB-D. Doktorant był autorem zasadniczych koncepcji, a także opracował oraz przygotował zasadniczą część artykułu.

W pracy opublikowanej na znaczącej konferencji IROS zaproponowano oraz eksperymentalnie przebadano oryginalne rozwiązania na potrzeby wyznaczania deskryptorów segmentów dla systemów SLAM pracujących w oparciu o LiDAR. W proponowanym podejściu odkrywane są deskryptory pomiarów LiDAR z uwzględnieniem wizualnego kontekstu segmentów. Nowa architektura głębokiej sieci neuronowej uczy się kontekstu wizualnego, który uzyskiwany jest z syntetycznych obrazów intensywności LiDAR. Proponowane podejście umożliwia tworzenie sygnatur lokalizacji o znaczącej sile dyskryminacyjnej. Metodę przebadano na dwóch ogólnie dostępnych repozytoriach danych. Badania eksperymentalne pokazały, że nowe deskryptory mają większą siłę dyskryminacyjną, a w szczególności umożliwiają bardziej niezawodne wykrywanie domknięcia pętli w SLAM. Doktorant był autorem koncepcji, a także przygotował zasadniczą część artykułu.

W autorskiej pracy, zamykającej cykl publikacji zaproponowano nowatorską metodę wykrywania segmentów płaszczyzn na obrazach pozyskiwanych z kamer w układzie stereoskopowym. Zaproponowano architekturę głębokiej sieci neuronowej, która rekonstruuje geometrię w oparciu o *cost volume* oraz tworzy deskryptory segmentów płaszczyzn. W celu

poprawy dokładności rekonstrukcji geometrii zaproponowano niezależną od parametrów kamery reprezentację wektorów normalnych. W trakcie wyznaczania pozy uwzględniana jest niepewność estymacji głębi, a także uwzględniane są granice segmentów płaszczyzn. Dzięki zaproponowanym deskryptorom do wyznaczenia dopasowania pobierana jest mniejsza liczba kandydatów w porównaniu do dopasowania wyznaczanego w oparciu o histogramy kolorów. Pokazano eksperymentalnie, że sieć neuronowa uzyskuje lepsze dokładności rekonstrukcji sceny niż konkurencyjne metody. Pokazano doświadczalnie, że na wymagającym repozytorium obrazów TERRINet, przy koniecznym warunku uzyskania bezbłędnych rozpoznań, system uzyskał lepsze wskaźniki rozpoznawania pozycji. Omawiana praca napisana jest bardzo starannie. Praca wskazuje, że Autor ma bardzo dobre wyczucie przedmiotowego problemu naukowego oraz dobrą intuicję odnośnie dalszych możliwych prac w tym obszarze tematycznym. Autor pokazał, że biegle opanował aparat matematyczny niezbędny do prowadzenia badań w tym obszarze.

Rozprawa wzbogaca naszą wiedzę, jest inspirująca i stanowi bardzo dobry materiał do dalszych prac. Rozwiązania zaproponowane w pracy są całościowe i mają nowatorski charakter. Autorskie propozycje rozwiązań wskazują na dobrą znajomość zagadnień związanych z przedmiotem rozprawy oraz na dobre wyczucie istoty prac o charakterze badawczym. Podjęty temat ma istotne walory praktyczne oraz poznawcze. Wyniki badań eksperymentalnych są spójne, przekonujące i należycie udokumentowane.

5. Słabych stron rozprawy jest niewiele i są one mniej istotne. Od strony poznawczej prace wzbogaciłoby odniesienie się do globalnej lokalizacji agenta w środowiskach dynamicznych. Z pewnością warto byłoby odnieść się szerzej do możliwości oraz ograniczeń technologii kamer w układzie stereoskopowym. Pracy nie zaszkodziłoby przeanalizowanie wrażliwości systemu na parametry, bardziej szczegółowe omówienie kluczowych parametrów systemu PlaneLoc2, w szczególności w kontekście wyników prezentowanych w Tab. 1. w pracy dotyczącej wspomnianego systemu. Interesujące byłoby przeanalizowanie krzepkości rozwiązania w szerszym kontekście, w tym także sytuacji lub zestawu parametrów, dla których odsetek „incorrect” byłby większy od zera.

Powyższe uwagi mają specyficzny charakter, gdyż nie umniejszają w najmniejszym stopniu wartościowych osiągnięć, a ukierunkowane są na uczynienie z bardzo dobrej rozprawy jeszcze lepszej.

6. W podsumowaniu stwierdzam, że recenzowana rozprawa mgr inż. Jana Wietrzykowskiego spełnia z wyraźnym nadmiarem wszystkie wymagania ustawowe i zwyczajowe stawiane rozprawom doktorskim, zarówno pod względem zawartości, jak i formy. Stwierdzam niniejszym, że przedstawiona mi do recenzji rozprawa doktorska spełnia kryteria określone w art. 13 ust. 1 Ustawy. Stąd wnoszę o dopuszczenie Pana Wietrzykowskiego do dalszych czynności przewodu doktorskiego. Stwierdzam, że przedstawiona do recenzji rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego, potwierdzając wiedzę oraz umiejętność samodzielnego prowadzenia badań naukowych przez Doktoranta. Podjęta tematyka badawcza jest aktualna i ma potencjał aplikacyjny. Ponadto, wysoki poziom merytoryczny i potencjał praktyczny zaprojektowanych rozwiązań oraz opublikowanie wyników badań w uznanych czasopiśmie z listy JCR, są przesłanką do wnioskowania do Rady Dyscypliny Automatyki, Elektroniki i Elektrotechniki Politechniki Poznańskiej o wyróżnienie rozprawy doktorskiej Pana mgr. inż. Jana Wietrzykowskiego.

Boydwan Kwoltek