

PRZEWODNICZĄCY RADY DYSCYPLINY  
Automatyka, Elektronika i Elektrotechnika

prof. dr hab. inż. Wojciech Szeląg

Warszawa, 23.08.2022

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

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

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

**Autor rozprawy:** Jan Wietrzykowski

### 1 Analiza strony merytorycznej rozprawy

#### 1.1 Obszar problemowy

Recenzowana rozprawa doktorska dotyczy zagadnień związanych z globalną lokalizacją kamery wewnątrz obiektów na podstawie obserwowanych danych, w połączeniu z budową mapy otoczenia (*ang.* simultaneous localization and mapping - SLAM). Główną osią rozprawy jest rozwiązanie problemów związanych z dotychczasowo stosowanymi metodami SLAM, w szczególności zmierzenie się z problemem zależności tych metod od jakości cech obiektów wykorzystywanych do lokalizacji. Ponieważ aktualnie stosowane metody w dużej mierze polegają na wizualnych cechach opisujących otoczenie, ich nierównomierne rozłożenie, wynikające np. z braku tekstury, może prowadzić do niezadowolających wyników lokalizacji. Trudności te są naturalne dla wielu aplikacji z zakresu robotyki oraz automatyki, jak również informatyki technicznej, a jednocześnie relatywnie mało alternatywnych form budowania reprezentacji otaczających obiektów zostało zaproponowane w literaturze zgłębiającej systemy typu SLAM. Uważam więc, że proponowana problematyka pracy jest nie tylko aktualna, ale również 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 lokalne, częściowe i niepewne wskazówki pochodzące z analizy płaszczyzn otaczających kamerę mogą wystarczyć do zbudowania probabilistycznej funkcji określającej położenie kamery, w szczególności w środowisku stworzonym przez człowieka, gdzie takie płaszczyzny są często spotykane. Dodatkowo, niejako dopełniając główną tezę pracy, Doktorant formułuje pomniejsze hipotezy badawcze dotyczące możliwości wykorzystania lokalnych geometrycznych cech do stworzenia globalnej funkcji prawdopodobieństwa pozycji kamery, wykorzystania płaszczyzn do wyliczenia pozy, jak również wykorzystania algorytmów uczenia głębokiego do identyfikacji i segmentacji płaszczyzn. Co prawda powyższe hipotezy nawiązują do głównej tezy pracy, natomiast nie wyczerpują one wszystkich możliwych ścieżek badania głównej hipotezy pracy, pozostawiając Autorowi pole do dalszych badań nad analizowanym zagadnieniem. Drobnym mankamentem proponowanej ścieżki badawczej jest brak podjęcia tematu generalizacji proponowanych metod na przestrzenie zewnętrzne (*outdoor SLAM*). Brak takiej analizy powinien w ogólności zostać wskazany jako potencjalne ograniczenie proponowanych metod i przeanalizowany pod kątem sposobów radzenia sobie z sytuacjami, kiedy operacja lokalizacji następuje np. na krawędzi świata wewnętrznego i zewnętrznego.

Jako cel pracy Autor stawia rozwiązanie problemu lokalizacji na bazie niejednoznacznych cech wizualnych obiektów oraz stworzenie środowiska pozwalającego na zidentyfikowanie pozy obiektu na bazie analizy probabilistycznej, która wykorzystuje obszary planarne. 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.

Pierwszy rozdział rozprawy to wstęp, który opisuje motywację podjęcia pracy badawczej w określonym wyżej kierunku, krótko omawia aktualnie istniejące rozwiązania oraz ich ograniczenia. W rozdziale tym Autor opisuje również wkład w rozwój nauki opisany w dalszej części rozprawy, jak również stawiane hipotezy badawcze. Rozdział pierwszy kończy się krótkim streszczeniem zestawu publikacji, będących główną częścią merytoryczną pracy, a znajdujących się jako załączniki w rozdziale drugim. Ten swoisty autoreferat, podsumowujący zaproponowane metody, stanowi syntezę uzyskanych przez Doktoranta rezultatów, jak również przewodnik po publikacjach pokazujący powiązania pomiędzy nimi oraz spójność tematyczną rzeszonych prac. Co prawda konstrukcja rozprawy polegająca na połączeniu w jednym rozdziale wszystkich publikacji, nie zmieniając ich formatowania w stosunku do oryginalnie opublikowanych prac, utrudnia odpowiednie zredagowanie wstępu, natomiast uważam, że przegląd literatury powiązanej z prowadzonymi badaniami mógłby bardziej przekrojowo opisywać analizowane tematy, w tym m.in. bardziej bezpośrednio odnosić się do obrazowych cech obiektów wykorzystywanych w systemach SLAM. Rozdział ten mógłby również poglądowo omówić kluczowy dla całej pracy konwencjonalny system SLAM, co pomogłoby czytelnikom przebrnąć przez wymienione w rozdziale drugim publikacje, a całą rozprawę uczyniłoby bardziej samowystarczalną. Wybrana forma prezentacji ułatwia czytelnikowi poruszanie się pomiędzy publikacjami, natomiast generuje problem redundancji pewnych części tekstu, takich jak wstępy czy analizy literaturowe artykułów. Jednocześnie należy przyznać, że taki styl pisania rozpraw doktorskich jest powszechnie akceptowany i należy obecnie do kanonu rozpraw doktorskich w dziedzinach informatyki i robotyki.

Przedstawione w rozdziale drugim publikacje opisują następujące proponowane przez Autora metody i podejścia do rozwiązania stawianego w rozprawie problemu:

1. wykorzystanie dwóch sposobów reprezentowania płaszczyzn, jednej bazującej na kwaternionach i drugiej bazującej na  $SE(3)$ , do narzucenia ograniczeń przy obliczaniu pozycji kamery. Reprezentacje te zostały przetestowane w środowisku  $g^2o$ , a przedstawione rezultaty wykazały, iż reprezentacja minimalna, bazująca na kwaternionach, daje nieznacznie lepsze rezultaty,
2. probabilistyczne środowisko do globalnej lokalizacji bazujące na dużej liczbie małych zbiorów geometrycznych cech, mające na celu stworzenie funkcji gęstości prawdopodobieństwa lokalizacji kamery,
3. uogólnienie zaproponowanego środowiska, nazwane PlaneLoc, pozwalające na unifikację modułu tworzącego lokalną i globalną mapę przestrzeni, które uniezależnia proponowane rozwiązanie od zewnętrznych bibliotek,
4. wykorzystanie głębokich sieci neuronowych do segmentacji i detekcji płaszczyzn, pozwalające na uniezależnienie rozwiązania od sensorów głębi,
5. opracowanie deskryptora obrazu bazującego na intensywności odczytów chmur punktów przez LIDAR,
6. rozszerzenie środowiska PlaneLoc o opisane powyżej modyfikacje, koronujące prace przedstawione w całej rozprawie.

Rozprawę kończy rozdział trzeci, w którym Autor przedstawia wnioski końcowe, podsumowuje przedstawione w rozdziale drugim publikacje, jak również wskazuje główne obszary badawcze, na które publikacje te mają wpływ i możliwe dalsze prace nad proponowanymi metodami.

Rozprawa zawiera reprezentatywną dla poruszanej problematyki bibliografię, która w większości przypadków właściwie ilustruje omawiane zagadnienia i dokumentuje wkład własny Autora.

## 1.2 Ocena wyników oraz stopnia ich oryginalności

Recenzowana rozprawa dotyczy rozwiązania problemów występujących w systemach lokalizacji agenta bazujących na deskryptorach obrazu. Oryginalne wyniki osiągnięte zostały przez Doktoranta zarówno w kontekście metodologii podejścia do problemu lokalizacji (wykorzystanie płaszczyzn i segmentów jako źródeł cech umożliwiających lokalizację), jak i poszczególnych jego modułów (tworzenie efektywnych reprezentacji płaszczyzn i ich detekcji przy wykorzystaniu uczenia głębokiego).

Zaprezentowane w rozprawie rozwiązania postawionego problemu badawczego bazują na nowych pomysłach lub zawierają wyraźne modyfikacje w stosunku do metod przedstawionych w literaturze.

Na szczególną uwagę zasługuje konsekwencja w dążeniu do opracowania rozwiązania, które eliminuje ograniczenia istniejących metod, pozwalająca na zastosowanie proponowanego finalnego środowiska PlaneLoc2 w rzeczywistych aplikacjach.

Wśród wielu przedstawionych w rozprawie rozwiązań, dodatków technicznych i implementacyjnych oraz usprawnień istniejących metod należy wyróżnić:

- nowatorskie metody wykorzystania reprezentacji płaszczyzn do ograniczania optymalizacji, mające na celu bardziej efektywne poszukiwanie lokalizacji kamery,
- nowe probabilistyczne środowisko lokalizacji PlaneLoc wykorzystujące wiele mniejszych zbiorów cech geometrycznych do zbudowania funkcji gęstości prawdopodobieństwa lokalizacji kamery, na podstawie trójek par segmentów,
- wprowadzenie nowego zbioru danych do lokalizacji wizualnej PUT RGB-D/WORKSHOP dataset,
- generalizację proponowanego systemu pozwalającą na połączenie modułów budujących lokalną i globalną mapę, bazującą na segmentach płaszczyzn,
- zaproponowanie nowej miary odległości pomiędzy segmentami, która minimalizuje problemy związane z parametryzacją płaszczyzn,
- zaprojektowanie nowej architektury głębokiej sieci neuronowej do detekcji płaszczyzn, która pozwala na uniezależnienie systemu od dokładności i zasięgu sensorów głębi,
- wprowadzenie syntetycznego zbioru SceneNet Stereo pozwalającego na wytrenowanie sieci neuronowej do detekcji płaszczyzn,
- opracowanie deskryptora wykorzystującego intensywność pomiarów LIDER w chmurach punktów 3D,
- spójne połączenie proponowanych metod i modyfikacji w jedno funkcjonalne środowisko PlaneLoc2, oferujące znaczną przewagę nad konkurencyjnymi rozwiązaniami, takimi jak ORB-SLAM3.

Należy podkreślić, że pomimo tego, iż część publikacji wchodzących w skład przedstawionej rozprawy ma kilku autorów, wkład Doktoranta jest jasno opisany i stanowi dominującą część prac, o czym świadczy również jego pierwsza pozycja na liście autorów.

Zauważone 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 artykule [Wietrzykowski 2016] brakuje empirycznego porównania z innymi metodami reprezentacji płaszczyzn ponad te proponowane przez Doktoranta. Czy Doktorant prowadził badania mające na celu uzyskanie wyników dla standardowych trójwymiarowych reprezentacji płaszczyzn? Jaka jest ich relacja w stosunku do proponowanych reprezentacji płaszczyzn?
- W artykule [Wietrzykowski 2016] brakuje eksperymentów przeprowadzonych na rzeczywistych danych, gdyż zaprezentowane rezultaty pokazują wyłącznie ewaluacje na danych syntetycznych. Czy Autor może przedstawić wyniki na zbiorach rzeczywistych danych?
- Sekcja 6 artykułu [Wietrzykowski 2016] nie zawiera szczegółowego opisu dotyczącego pomiarów czasu działania systemu wykorzystującego obie reprezentacje płaszczyzn - czy zostały przeprowadzone wielokrotne próby uruchomienia systemu czy raportowane wyniki pochodzą z jednego przebiegu?
- W artykule [Wietrzykowski and Skrzypczyński 2017], zaproponowana metoda wyboru tripletów wydaje się dosyć uproszczona, wykorzystująca w szczególności różnice w histogramach budowanych na przestrzeni barw HSV. Wybór tej przestrzeni barw, podobnie jak założenie, że liczba

par musi być ograniczona, nie jest uzasadniony w przedstawionym artykule i wymaga dodatkowej analizy, w tym m.in. eksperymentów potwierdzających efektywność tych propozycji, np. w stosunku do innych przestrzeni barw (HSL, YUV, RGB) czy ich reprezentacji (parametry histogramu).

- W artykule [Wietrzykowski and Skrzypczyński 2019] brakuje eksperymentalnych wyników uzasadniających wybór parametrów metody, w szczególności takich jak  $\tau_{div}$  czy  $\tau_{met}$ . Co prawda wybiórcze wyniki ustawień kilku parametrów są pokazane w Tabeli 1, natomiast analiza wpływu tych parametrów na funkcjonowanie całego systemu jest pominięta. Czy doktorant może przedstawić wyniki badań prowadzące do wskazanych w artykule parametrów?
- W artykule [Wietrzykowski and Belter 2022] brakuje opisu szczegółów implementacyjnych dotyczących weryfikacji prędkości inferencji - w artykule podana jest średnia wartość 0.419s, która jest ok. dwukrotnie dłuższa niż operacja wykonana dla metody Plane R-CNN, jednak brakuje informacji dot. liczby powtórzeń operacji oraz zbioru, na którym zostały one wykonane.
- W artykule [Wietrzykowski 2022] rezultaty ewaluacji wydajności algorytmu przedstawione są bardzo pobieżnie i wyłącznie w sekcji III, bez odniesienia do rezultatów innych konkurujących metod.
- W artykule [Wietrzykowski 2022] w sekcji VI.B. opis metody ustawienia parametrów jest nadmiernie krótki i de facto sprowadzony do stwierdzenia, że parametry są ustawione na podstawie analizy pierwszej sceny. Potrzebne jest bardziej szczegółowe określenie jak procedura ta wygląda, jakie są jej przewagi w stosunku do wyznaczenia parametrów w sposób globalny i przede wszystkim jak kosztowna obliczeniowo jest.
- Generalna uwaga dotycząca ewaluacji proponowanych w rozprawie algorytmów dotyczy możliwości zbiorczego przeanalizowania wpływu poszczególnych elementów na finalne rezultaty uzyskiwane przez system (*ablation study*). Rozprawa byłaby dużo pełniejsza, gdyby Doktorant przedstawił w ramach podsumowania wyniki eksperymentalne pokazujące jak poszczególne modyfikacje wpływają na końcowe działanie systemu, zmierzone przy użyciu standardowych miar błędu i precyzji, w szczególności wykorzystując publicznie dostępne rzeczywiste zbiory danych. Warto byłoby takie porównanie odnieść do istniejących systemów, co częściowo zostało uczynione w pracy [Wietrzykowski 2022], jednak analiza ta nie jest w moim odczuciu kompletna w kontekście wszystkich proponowanych w rozprawie usprawnień i modyfikacji. W szczególności istotne byłoby przedstawienie takiej ewaluacji, wykorzystując realistyczne dane *robotyczne*, tzn. pochodzące z sensorów umieszczonych na robotach, co w pełni uzasadniałoby wskazane w tytule rozprawy zastosowanie proponowanych metod w robotyce.

### 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.

- Postawiony w rozprawie problem badawczy został zawężony do lokalizacji wewnątrz budynków, chociaż niektóre proponowane metody, w tym np. deskryptor bazujący na intensywności odczytów LIDARa [Wietrzykowski and Skrzypczyński 2021], został z powodzeniem zastosowany do zbioru danych KITTI, zawierającego obrazy spoza budynków. Jakie modyfikacje proponowanych metod bazujących na płaszczyznach mogłyby zostać zaproponowane, żeby prezentowane środowiska PlaneLoc i PlaneLoc2 mogłyby być stosowane również w takim środowisku zewnętrznym?
- W artykule [Wietrzykowski and Skrzypczyński 2017] wybór konkretnych tripletów par segmentów do stworzenia funkcji gęstości prawdopodobieństwa wydaje się dosyć arbitralny. Standardowe metody uczenia głębokiego, które również wykorzystują triplety np. do budowania reprezentacji obrazów, posilkują się dodatkowymi mechanizmami wyboru tripletów, które bazują swoje decyzje na tym jak informatywne dla systemu jest dodanie kolejnego tripletu. Przykładem takiej metody jest *hard negative mining* [Jin et al. ECCV'2018]. W artykule wskazane jest, że triplety tworzone są z wizualnie podobnych płaszczyzn, a następnie ważone. Czy taka decyzja została właśnie podyktowana chęcią wytworzenia "trudnych" przypadków dla uzyskania lepszych efektów czy istniały inne motywacje?



## 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, trzy rozdziały zasadnicze (w tym "Wstęp" oraz "Wnioski"), a także bibliografię. Praca liczy 89 stron.

Bibliografia liczy 33 uporządkowane alfabetycznie pozycje (nie wliczane są tutaj pozycje bibliograficzne z przedstawionych w rozdziale drugim artykułów). W rozprawie cytowanych jest 6 prac własnych Autora (w tym 4 współautorskie). Prace te, będące jednocześnie zawartością rozdziału drugiego, pochodzą z materiałów konferencyjnych jak również z renomowanych czasopism branżowych. Spośród znajdujących się w bibliografii prac praktycznie wszystkie ukazały się w przeciągu ostatnich 5 lat co świadczy o dobrym umiejscowieniu tematyki rozprawy w aktualnym w skali światowej nurcie badań. Bibliografia rozprawy, po włączeniu w nią utworów cytowanych w ramach załączonych artykułów, nie budzi zastrzeżeń od strony merytorycznej, a jej redakcja jest staranna i nie dostrzegłem w niej żadnych błędów.

Układ rozprawy jest prawidłowy, jest ona także starannie opracowana pod względem edytorskim, w szczególności rozdziały pierwszy i trzeci są odpowiednio dopracowane i nie widać w nich 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

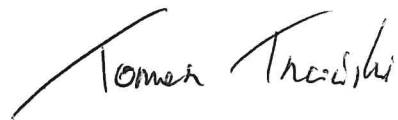
Tekst rozprawy jest w przeważającej większości poprawny pod względem językowym i stylistycznym. Podczas lektury zauważyłem nieliczne błędy redakcyjne i pomyłki wymagające korekty, np:

- str. 3, w. 4: "renowned journals [...] top-tier conferences..." - to dosyć subiektywna ocena, której praca naukowa powinna unikać, pozostawiając ją w gestii recenzentów czy czytelników.
- str. 3, w. 7: "making it reliable and robust is of utmost importance" - zbędne sformułowanie, dodatkowo bazujące na subiektywnej opinii.
- str. 3, w. 17: "As a results": powinno być "As a result".
- str. 4, w. 2: Brakuje spacji przed "Omówiono".
- str. 4, w. 5: "renomowanych [...] uznanych" - uwaga podobna do komentarza odnoszącego się do wersji angielskiej.
- str. 16, w. 2: "adjacency of considered views" - powinno być "where there is an adjacency of considered views", lub podobnie. Co więcej, warto tutaj dodać referencję potwierdzającą zawarte w zdaniu twierdzenie.
- str. 16, w. 9 od dołu: "using a hierarchical approach, where features are matched [...]" - to zdanie nie jest prawdziwe, tzn. również nieuczzone deskryptory (np. SIFT) wykorzystują hierarchię bazującą na wielu skalach rozdzielczości obrazu.
- str. 26 2. paragraf po prawej: "features that minimizes" - powinno być "that minimize"
- str. 29 lewa kolumna: "it's" - powinno być "its"
- str. 63 prawa kolumna: "that researched" - powinno być "that carried out research on"
- str. 64 prawa kolumna: "that are merged down-stream the computations" - niejasne sformułowanie, powinno być "that are merged down the processing pipeline", lub podobnie.
- str. 65 lewa kolumna: "tells the DNN" - niepotrzebna personifikacja, lepiej: "indicates to the DNN".
- str. 65 lewa kolumna: "Typically to image processing DNNs" - powinno być: "Typically for the DNNs that process images".
- str. 78 prawa kolumna: "outperforms the classic" - powinno być outperforms the classic one.

### 3 Konkluzja

Uważam, że recenzowana dysertacja Pana mgr. inż. Jana Wietrzykowskiego **spełnia z nadmiarem 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 oryginalne koncepcje, rozwiązania istotnych problemów w dziedzinach informatyki i robotyki. Rezultaty nawiązują do aktualnego stanu wiedzy i mają znaczenie praktyczne, czego potwierdzeniem są liczne publikacje w renomowanych czasopismach (m.in. IEEE Robotics and Automation Letters, IF: 4.321, 200 pkt. MEiN) oraz, przede wszystkim, na najważniejszych konferencjach międzynarodowych z dziedziny robotyki (IROS, CORE A, 140 pkt. MEiN). 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 bezapelacyjnie pozytywną ocenę oryginalności i poziomu merytorycznego przedstawionej rozprawy.

Sformułowanie problemu badawczego, zaproponowane oryginalne metody jego rozwiązania, sposób przeprowadzenia badań oraz zademonstrowana umiejętność formułowania wniosków świadczą o przygotowaniu Doktoranta do pracy naukowej. Na tej podstawie **wnioskuję o dopuszczenie mgr. inż. Jana Wietrzykowskiego do dalszych kroków procedury uzyskania stopnia doktora nauk technicznych**. Co więcej, ze względu na wysoką jakość przedstawionych publikacji oraz renomowane miejsca, w których się one ukazały, **wnioskuję o wyróżnienie pracy doktorskiej**.



Tomasz Trzciniński