

RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY WYDZIAŁU MECHATRONIKI POLITECHNIKI WARSZAWSKIEJ

Tytuł rozprawy: Rozpoznawanie obiektów w chmurach punktów
na potrzeby robotyki mobilnej

Autor rozprawy: mgr inż. Bogdan Harasymowicz-Boggio

1. Analiza strony merytorycznej rozprawy

1.1. Obszar problemowy

Recenzowana rozprawa doktorska dotyczy zagadnienia rozpoznawania obiektów na podstawie danych RGB-D w postaci chmur punktów. Kontekst aplikacyjny rozważanego w rozprawie zagadnienia dotyczy robotyki, a szczególnie potrzeby rozpoznawania typów obiektów oraz ich instancji (konkretnych egzemplarzy) przez mobilne roboty usługowe. Należy jednak zwrócić uwagę, że zagadnienie rozpoznawania obiektów stanowi część szerszej problematyki rozumienia sceny na podstawie obrazów (ang. *image understanding*). Problematyka ta jest jedną z ważniejszych we współczesnej informatyce, a jej znaczenie dla robotyki systematycznie rośnie wraz ze wzrostem oczekiwanego stopnia autonomii robotów, głównie mobilnych. Rozważane w rozprawie zagadnienie jest przedmiotem intensywnych badań w wielu ośrodkach naukowych na świecie, a także przedmiotem zainteresowania wielu firm komercyjnych. Jest to związane z potencjałem jaki może dać różnym urządzeniom i systemom technicznym (nie tylko robotom) zdolność do interpretacji obserwowanej sceny w kategoriach semantycznych. Biorąc pod uwagę aktualny stan badań naukowych w dziedzinie robotyki oraz widzenia maszynowego uważam problematykę rozprawy za aktualną i dającą możliwość uzyskania nowych wyników teoretycznych i aplikacyjnych.

W tezie postawionej w rozprawie Autor mówi o różnicach pomiędzy ogólnym zadaniem rozpoznawania obiektów, a rozpoznawaniem obiektów w robotyce, oraz korzyściach płynących z wykorzystania wiedzy o kontekście semantycznym sceny i ograniczeniach percepcyjnych sensora. Pierwsza część tezy jest bardzo ogólna, a więc trudno ją zweryfikować. Jednak część druga trafnie odnosi się do najbardziej oryginalnych rozwiązań przedstawionych w rozprawie.

Za główny cel pracy Autor obrał opracowanie zbioru metod rozwiązujących problem rozpoznawania obiektów występujących na scenach obserwowanych za pomocą sensora RGB-D (klasy Kinect), spełniających pewne założenia specyfikujące ten problem w kontekście robotów mobilnych. Związane z realizacją tego celu badania mieszczą się w obszarze robotyki oraz takich dziedzin informatyki jak widzenie maszynowe i uczenie maszynowe.

We wstępie (rozdział pierwszy) Autor przedstawia uzasadnienie podjęcia tematu badawczego i właściwie określa specyfikę zastosowań w robotyce jako obszar wymagający dalszych badań w zakresie rozpoznawania obiektów na podstawie obrazów. W odniesieniu do tego obszaru definiuje szereg założeń, dotyczących między innymi dopuszczalnego czasu przetwarzania informacji o obserwowanej scenie, uczenia bez wykorzystania masywnych zbiorów danych, stopnia uporządkowania obserwowanych scen, dostępności *a priori* wiedzy dotyczącej relacji topologicznych między obiektami oraz możliwości zbierania danych poprzez aktywną eksplorację sceny. W ten sposób zostaje zdefiniowany interesujący

wycinek bardzo szerokiego frontu badań, w którym Autor spodziewa się uzyskania nowych wyników dysponując dość ograniczonymi środkami, np. w zakresie zasobów obliczeniowych czy też dostępu do sekwencji uczących.

W rozdziale drugim na podstawie studiów literaturowych podjęto próbę nakreślenia stanu wiedzy w obszarze metod rozpoznawania obiektów na podstawie obrazów 2D i danych trójwymiarowych, wyraźnie koncentrując się na drugiej z tych grup rozwiązań. Rozdziały trzeci, czwarty i piąty dotyczą nowych rozwiązań w dziedzinie rozpoznawania obiektów na podstawie chmur punktów RGB-D i zawierają zasadnicze, oryginalne treści rozprawy. Rozdział trzeci poświęcono sposobom uwzględnienia kontekstu w rozpoznawaniu obiektów. Przez kontekst rozumiane są tutaj głównie typowe relacje topologiczne między obiektami (wsółwystępowanie, sąsiedztwo, wzajemne ułożenie w przestrzeni). W rozdziale czwartym przedstawiono rozwiązanie systemu rozpoznawania obiektów pozwalające na uwzględnienie niepewności oraz niewiedzy wynikających z ograniczeń systemu pomiarowego, takich jak niedokładność pomiaru odległości, jego ograniczony zasięg i wzajemne przesłonięcia obserwowanych obiektów. Rozdział piąty przedstawia natomiast proponowane przez Autora rozwiązania zmierzające do uzyskania możliwości analizy scen w czasie zbliżonym do rzeczywistego, a dotyczące heurystycznego przyspieszenia procesu rozpoznawania oraz masowo-równoległej metody segmentacji chmury punktów. W rozdziale szóstym dość ogólnie przedstawiono implementację systemu rozpoznawania obiektów dla autonomicznego robota mobilnego, wykorzystującą przedstawione w poprzednich rozdziałach metody i algorytmy. Rozprawę zamyka rozdział siódmy, prezentujący podsumowanie i wnioski wraz z wyszczególnieniem nowych, oryginalnych elementów. Rozprawa zawiera reprezentatywną dla poruszanej problematyki bibliografię, która w większości przypadków właściwie ilustruje omawiane zagadnienia oraz dokumentuje wkład własny Autora.

1.2. Ocena wyników oraz stopnia ich oryginalności

W recenzowanej rozprawie Autor podjął się złożonego zadania budowy wizyjnego systemu rozpoznawania obiektów spełniającego specyficzne założenia, pozwalające na jego integrację z systemem sterowania robota mobilnego. Osiągnięte oryginalne wyniki dotyczą czterech obszarów: rozpoznawania obiektów z wykorzystaniem kontekstu semantycznego, wykorzystania w rozpoznawaniu obiektów informacji o niepewności i niewiedzy wynikających z ograniczeń percepcyjnych, heurystycznego ograniczania przestrzeni poszukiwań metod rozpoznawania obiektów wykorzystujących deskryptory lokalne oraz masowo-równoległej segmentacji niezorganizowanych chmur punktów. W każdym z tych obszarów przedstawiono oryginalne koncepcje i nowe lub zmodyfikowane algorytmy oraz wyniki badań eksperymentalnych potwierdzające poprawność zaproponowanych rozwiązań. Wśród prezentowanych w rozprawie nowych koncepcji oraz nowych lub zmodyfikowanych algorytmów należy wyróżnić:

- wykorzystanie losowych pól Markowa do znajdowania najlepszej konfiguracji obiektów (nazywanej w rozprawie teorią sceny) wyjaśniającej uzyskane pomiary z uwzględnieniem kontekstu semantycznego wynikającego z możliwych relacji między obiektami;
- zastosowanie formalizmu Dempstera-Shafera do modelowania niepewności i niewiedzy wynikającej z niedoskonałości pomiarów RGB-D oraz nieuporządkowanego charakteru obserwowanej sceny;
- metodę heurystycznego ograniczania przestrzeni poszukiwań w zadaniu rozpoznawania obiektów z użyciem deskryptorów lokalnych poprzez kilkustopniową redukcję chmury punktów RGB-D;
- algorytm szybkiej segmentacji niezorganizowanych chmur punktów poprzez rozrost ziarna przeznaczony do implementacji masowo-równoległej z użyciem akceleratora GPGPU.
- koncepcja, realizacja i ewaluacja eksperymentalna systemu rozpoznawania obiektów na podstawie danych RGB-D przeznaczonego dla robota mobilnego.

Koncepcje przedstawione w rozprawie cechują się znacznym stopniem oryginalności. Autor w kolejnych rozdziałach stopniowo prezentuje rozwiązania pozwalające na spełnienie większości z przyjętych we wstępie założeń decydujących o specyfice rozpoznawania obiektów trójwymiarowych przez robota mobilnego. Poczynając od sposobu uwzględniania kontekstu semantycznego, poprzez uwzględnienie niepewności i niewiedzy, aż po techniki pozwalające na interpretację sceny w akceptowalnym z punktu widzenia użytkownika czasie. Autor swobodnie porusza się nie tylko w obszarze systemów wizyjnych RGB-D, lecz także w zakresie metod rozpoznawania wzorców i uczenia maszynowego, konsekwentnie wybierając techniki przetwarzania danych pozwalające zbudować system zgodny z założeniami. Niewątpliwie pomaga mu w tym wiedza z zakresu neurobiologii, gdyż uzasadniając strukturę i sposób działania proponowanych rozwiązań wielokrotnie odwołuje się do literatury z tego obszaru (np. str. 71 rozprawy).

Słabszą stroną rozprawy jest natomiast powierzchowność lub niekompletność opisów pewnych istotnych komponentów niektórych z proponowanych rozwiązań.

- W podrozdziale 3.3 nie podano struktury dwuwymiarowej mapy cech prostych, o której mowa na str. 37. W podrozdziale 3.2 opisano trzy rodzaje cech i nie jest jasne w jaki sposób są one zorganizowane w mapę (obraz) analizowany za pomocą detektora krawędzi. Na tej samej stronie zdefiniowano miarę służącą do porównywania histogramów cech powierzchni segmentu “do histogramu obliczonego dla części modelowych widoków obiektu”, nie określając jednak, w jaki sposób powstają owe histogramy części modelowych, ani czym są wspomniane części. Można się domyślać, że przygotowanie histogramów części jest elementem procedury uczenia systemu, w rozprawie nie określono jednak tej procedury.
- Opisując zależności dotyczące losowego pola Markowa, Autor stwierdza odnosząc się do wzoru (3.14), że wartości funkcji $f_{AB}(x_A, x_B)$ “dla każdej relacji są dostarczane bezpośrednio do systemu”. Procedura uzyskiwania tych wartości nie jest wyjaśniona, a w szczególności nie wiadomo jaki jest zakres dobieranych wartości oraz czy są one uniwersalne (specyficzne dla relacji danych obiektów), czy też specyficzne dla konkretnej sceny.
- W podrozdziale 3.5 do znalezienia rozwiązania równania (3.9) wykorzystano algorytm genetyczny. Autor komentuje wprawdzie pewne jego właściwości, zwracając uwagę na być może stanowiącą element oryginalny, procedurę optymalizacji lokalnej, jednak brak jest w rozprawie opisu całego algorytmu, np. w postaci schematu blokowego. Nie zostały także zdefiniowane operatory genetyczne, nie zdefiniowano sposobu kodowania osobników i nie podano jawnie funkcji celu. Algorytm genetyczny, jak można się domyślać ten sam, został użyty także w podrozdziale 4.3 do znalezienia konfiguracji etykiet obiektów najlepiej wyjaśniającej scenę w warunkach wnioskowania z niepewnością. Także w tym fragmencie rozprawy algorytm nie został poprawnie wyspecyfikowany.
- W podrozdziale 5.3 niejasna jest definicja i sposób wyznaczania punktów kotwiczących, w szczególności niezrozumiałe jest pojęcie “punktów o stosunkowo dużej częstotliwości”, które pojawia się na str. 80. W podrozdziale tym pełniejszego opisu wymagają też cechy D2, D3, D4, oraz podobnie jak wcześniej w rozdziale 3 sposób wyznaczania widoków modelowych używanych w teście korelacji Pearsona.

Brak również w recenzowanej pracy szczegółowego opisu danych używanych w eksperymentach. W niektórych przypadkach, np. Rys. 3.10 i Rys. 4.5 nie podano dla jakich danych (scen) uzyskano prezentowane wyniki ilościowe. W rozdziale 6, dotyczącym implementacji systemu na pokładzie robota mobilnego, nie w każdym przypadku oczywiste jest, które wersje opisywanych rozwiązań zostały użyte w kompletnym systemie poddawanych badaniom eksperymentalnym. Niemniej jednak wykonane badania pozwoliły na jakościową oraz przynajmniej częściowo ilościową ocenę osiągniętych wyników. Wnioski wyciągnięte przez Autora na podstawie tych wyników wskazują na to, że cel podjętych badań został osiągnięty.

1.3. Uwagi szczegółowe

W poniższej części recenzji odnoszę się szczegółowo do kolejnych rozdziałów manuskryptu zwracając uwagę na dostrzeżone usterki merytoryczne, oraz zadając pytania, na które oczekuję odpowiedzi Autora, podobnie jak na uwagi ogólniejszej natury przedstawione w poprzednim podrozdziale.

1. Str. 26, w. 10 od góry: nie rozwinięto akronimów wzmiankowanych deskryptorów.
2. Str. 38, wzór (3.6): jak podczas znajdowania transformacji maksymalizującej (3.6) uwzględniane jest założenie, że obiekty mogą obracać się tylko wokół osi pionowej?
3. Str. 40, w. 11 od góry: w jaki sposób scena dzielona jest na klastry segmentów?
4. Str. 41: wprowadzając dyskretne losowe pole Markowa i proponując jego modyfikację Autor nie wspomina o znanym z literatury rozwiązaniu jakim są *Conditional Random Fields* (CRF), używane między innymi w zadaniu klasyfikacji, np. S. Laible, Y. Khan, A. Zell, Terrain classification with conditional random fields on fused 3D lidar and camera data, Proc. ECMR, Barcelona, 2013.
5. Str. 38, wzór (4.10): czy pole powierzchni segmentu A_S nie zależy od odległości między sensorem a obserwowanym obiektem?
6. Str. 76, w. 9 od góry: co oznacza określenie “20-milimetrowe sąsiedztwo” punktu (jaki jest jego kształt i rzeczywisty rozmiar)?
7. Str. 77, Rys. 5.1: nie opisano znaczenia zmiennych K oraz L występujących na tym rysunku.

1.4. Zagadnienia dyskusyjne

Poniższe uwagi dotyczą ogólniejszych kwestii poruszonych w rozprawie i nie odnoszą się bezpośrednio do treści manuskryptu. Oczekuję jednak komentarza Autora dotyczącego tych uwag.

1. Jednym z oryginalnych elementów rozprawy jest wykorzystanie teorii Dempstera-Shafera (D-S) w klasyfikacji obiektów z uwzględnieniem niepewności i niewiedzy. Z innych badań, między innymi na polu robotyki mobilnej, a dotyczących map rastrowych wiadomo jednak, że podobne własności można uzyskać także wykorzystując inne formalizmy, np. zbiory rozmyte. Czy zdaniem Autora możliwe jest zastąpienie w proponowanym systemie formalizmu D-S formalizmem zbiorów rozmytych (lub innym podobnym) w celu uniknięcia pewnych problemów związanych np. z działaniem reguły agregacji Dempstera (wspomniano o nich na str. 63 rozprawy)?
2. W części dotyczącej eksperymentów z użyciem robota mobilnego (rozdział 6) wykorzystywane chmury punktów RGB-D powstają przez łączenie kolejnych pomiarów z sensora Kinect za pomocą programu KinectFusion. Nie jest całkowicie jasne, czy podczas agregacji danych robot pozostaje w spoczynku, czy też może się poruszać. Zapewne akwizycja danych w ruchu pozwala na rozszerzenie pola widzenia i ogólną redukcję czasu wykonania zadania. Czy istnieje jednak możliwość uwzględnienia w proponowanym systemie niepewności pomiarów wynikającej z niepewności pozycji robota?

2. Analiza strony formalnej rozprawy

2.1. Ocena układu pracy i redakcji manuskryptu

Recenzowana rozprawa doktorska obejmuje w kolejności: streszczenie w języku polskim, streszczenie w języku angielskim, spis treści, listę użytych symboli, siedem rozdziałów zasadniczych (w tym “Wstęp” oraz “Podsumowanie i wnioski”) oraz bibliografię. Praca liczy 111 ponumerowanych stron.

Bibliografia składa się ze 115 pozycji, w tym 105 publikacji oraz 10 stron internetowych. W rozprawie cytowanych jest 7 prac własnych Autora (współautorskich) oraz dwa źródła internetowe, których jest on współautorem. Ponad 60 z cytowanych prac ukazało się po roku 2010, a ponad 30 w ciągu ostatnich trzech lat, co świadczy o umiejscowieniu tematyki rozprawy w aktualnym nurcie badań światowych. Bibliografia nie budzi zastrzeżeń od strony merytorycznej, a jej redakcja jest staranna. Wskazać można jedynie pojedyncze drobne błędy redakcyjne ([48],[81],[106-115]).

Układ rozprawy jest prawidłowy, jest ona także starannie opracowana pod względem edytorskim oraz pod względem ilustracyjnym. Użyta terminologia jest właściwa dla obszaru problemowego rozprawy, zarówno w zakresie robotyki jak i sztucznej inteligencji oraz przetwarzania obrazu.

2.2. Uwagi szczegółowe

Tekst rozprawy jest poprawny pod względem językowym i stylistycznym. Podczas lektury manuskryptu zauważyłem jednak nieliczne błędy redakcyjne lub pomyłki, np.:

1. str. 23, w. 15 od góry: “mogą być stosowanie”,
2. str. 27, w. 2 od góry: “to polega to na”,
3. str. 32, w. 7 od góry: “będących bardzo podobnych”,
4. str. 34, Rys. 3.1: w podpisie mowa jest o ośmiokątach, podczas gdy na rysunku widzimy sześciokąty,
5. str. 42, Rys. 3.7: w podpisie “Makrowa”,
6. str. 44, w. 10 od dołu: “jednak jakoś klasyfikacji”,
7. str. 44, w. 9 od dołu: “przekładając wiedzę”,
8. str. 57, w. 7 od dołu: “odpowiada za zgromadzoną możliwość alternatywnych hipotez” (frazą niezrozumiałą),

3. Konkluzja

Uważam, że recenzowana dysertacja Pana mgr inż. Bogdana Harasymowicza-Boggio spełnia wymogi stawiane rozprawom doktorskim przez *Ustawę o stopniach naukowych i tytułach naukowych* z 14 marca 2003 roku (Dz. U. Nr 65 z 2003 r.), ponieważ zawiera oryginalną koncepcję rozwiązania istotnych problemów w dziedzinie robotyki oraz informatyki. Rezultaty nawiązują do aktualnego stanu badań i mają znaczenie praktyczne. Uzyskane wyniki eksperymentalne w wystarczającym stopniu dokumentują poprawność proponowanych rozwiązań koncepcyjnych i algorytmicznych oraz skuteczność działania ich implementacji. Uwagi krytyczne sformułowane w treści recenzji, po części mające charakter dyskusyjny, nie zmniejszają mojej pozytywnej oceny oryginalności i poziomu merytorycznego rozprawy.

Sformułowanie problemu badawczego, zaproponowane oryginalne metody i algorytmy 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. Bogdana Harasymowicza-Boggio do publicznej obrony.



Piotr Skrzypczyński