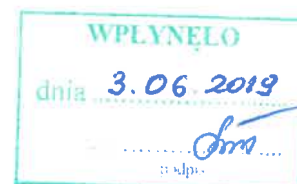




Dr hab. Grażyna Ślusarczyk
Zakład Projektowania i Grafiki Komputerowej
Wydział Fizyki, Astronomii i Informatyki Stosowanej
Uniwersytet Jagielloński

Kraków, 28. 05. 2019

RECENZJA ROZPRAWY DOKTORSKIEJ



Tytuł rozprawy: System wizyjny dla autonomicznego robota poruszającego się w trzech wymiarach

Autor rozprawy: mgr Piotr Śmigielski

Promotor rozprawy: dr hab. Andrzej Bielecki, prof. AGH

I. Informacje ogólne

Oceniana rozprawa doktorska mgr Piotra Śmigielskiego zatytułowana „System wizyjny dla autonomicznego robota poruszającego się w trzech wymiarach” stanowi opracowanie o objętości 113 stron, zawierające 84 rysunki oraz spis literatury obejmujący 92 pozycje. Autorskie algorytmy dostępne są pod adresami:

<https://github.com/smigielp/SceneAnalysisLib> i

<https://github.com/smigielp/SceneAnalysisTests>.

II. O przedmiocie rozprawy

Rozprawa pana mgr Piotra Śmigielskiego dotyczy opracowania systemu wizyjnego dla robota wyposażonego w jedną kamerę, który zapewniłby mu autonomię działania w oparciu o analizę sceny.

Autonomiczne roboty mobilne odgrywają i będą w przyszłości odgrywać znaczną rolę zarówno w zastosowaniach przemysłowych, transportowych, wojskowych, jak i eksploracyjnych. Zautomatyzowane systemy kontroli wizyjnej wykorzystujące algorytmy przetwarzania i rozpoznawania obrazów umożliwiają obserwację procesów produkcyjnych, ich szybką i niezawodną inspekcję, a także pozwalają kontrolować wybrane obszary i rejestrować określone zjawiska. Możliwości dostępnych dzisiaj systemów wizyjnych pozwalają na ich integrację z robotami przemysłowymi oraz wykorzystanie do realizacji złożonych algorytmów sterowania w oparciu o informacje o stanie sceny, w której roboty te pracują. Systemy wizyjne umożliwiające rejestrację i analizę obrazu wykorzystywane są w bezałogowych pojazdach rozpoznawczych - dronach powietrznych i naziemnych, a także nawodnych i podwodnych jednostkach pływających. Odpowiednie metody wykrywania i klasyfikacji obiektów obrazu pozwalają

skuteczne sterować robotem w różnych warunkach atmosferycznych i przy zmieniającym się oświetleniu.

Jednakże we współcześnie konstruowanych robotach mobilnych, pomimo ogromnych postępów widocznych w tej dziedzinie, efektywne wykorzystanie obrazu nadal znacząco odbiega od oczekiwań, szczególnie w przypadku robotów projektowanych z myślą o interakcji z człowiekiem. Wynika to m.in. z niedoskonałego sposobu pozyskiwania i przetwarzania informacji o częstokroć niedeterministycznym otoczeniu. Dlatego potrzebne są coraz bardziej zaawansowane algorytmy przetwarzania obrazu umożliwiające robotowi pracę w niezdefiniowanym wcześniej środowisku, a rozwój i implementacja trójwymiarowego postrzegania rzeczywistości stanowi jeden z kierunków rozwojowych robotyki mobilnej.

W rozprawie Autor zajmuje się algorytmami przetwarzania obrazu, rozpoznawania obiektów, konstruowaniem trójwymiarowego modelu elementów sceny, tworzeniem reprezentacji sceny i jej analizą. Autor rozprawy opracował moduł kognitywny, który umożliwia rozumienie sceny na podstawie jej grafowego modelu uwzględniającego kontekst przestrzenny znajdujących się na niej obiektów. Omówiony został także moduł nawigacji robota umożliwiający mu poruszanie się w obrębie sceny i budowę modelu sceny. Działanie przedstawionych algorytmów zostało zweryfikowane z wykorzystaniem symulowanego robota latającego operującego w sztucznym środowisku testowym. W ramach pracy skonstruowany został również prawdziwy robot latający umożliwiający weryfikację działania opracowanego systemu wizyjnego w plenerze.

III. O postawionym w rozprawie celu

Cel jaki postawił sobie w rozprawie doktorant to zaproponowanie i zaimplementowanie zestawu efektywnych algorytmów strukturalnej analizy sceny składających się na działający w czasie rzeczywistym system wizyjny autonomicznego robota latającego wyposażonego w jedną kamerę. Zastosowana została metoda wektoryzacji obrazu, metoda strukturalnego rozpoznawania obiektów sceny oraz algorytm konstrukcji trójwymiarowych modeli obiektów na podstawie rzutów ich ścian wykonanych z różnych kierunków. Zaimplementowane algorytmy rozpoznawania elementów sceny i postrzegania ich wzajemnej lokalizacji bazują na stworzonym tu grafowym modelu sceny i umożliwiają bezkolizyjną nawigację w jej obszarze.

Cel postawiony przez Autora należy do podstawowych zagadnień z dziedziny robotów mobilnych, gdyż jak wiadomo warunkiem koniecznym uzyskania autonomii robota jest skuteczna percepcja cech otoczenia i określenie przestrzennych relacji między obiektami sceny.

IV. Przegląd zawartości rozprawy

Rozdział 1 rozważa typowe zadania przed jakimi stawiane są roboty autonomiczne i podkreśla znaczenie systemów wizyjnych dla robotów mobilnych zarówno w kontekście analizy sceny jak i lokalizacji w przestrzeni. Omawiane są ograniczenia nałożone na systemy wizyjne dla robotów latających związane z niewielkim udźwigniem takich maszyn, co wiąże się z koniecznością stosowania jednostek obliczeniowych o niewielkich rozmiarach a co za tym idzie algorytmów o niewielkiej złożoności obliczeniowej. Zaznaczona została także potrzeba weryfikacji działania proponowanych rozwiązań w symulowanym środowisku. Następnie Autor przedstawia cel, który pragnie zrealizować w

rozprawie. Na zakończenie rozdziału przedstawiony i omówiony jest schemat logiczny procesu przetwarzania obrazu realizowanego przez zaproponowany w pracy system wizyjny.

Rozdział 2 prezentuje aktualny stan badań dotyczący rozpoznawania obrazów i analizy sceny w kontekście robotyki. Możliwości robotów, ich konstrukcja i funkcjonalności modułów kontrolnych zależą od misji do jakich są one przeznaczone. Rozwój badań w dziedzinie automatyki, elektroniki oraz oprogramowania pozwala na realizację coraz bardziej złożonych zadań. Autor omawia różne źródła informacji o otoczeniu wykorzystywane w robotyce, ze szczególnym uwzględnieniem sensorów wizyjnych. Następnie przedstawiony jest problem uzyskiwania autonomii przez łaziki marsjańskie za pomocą systemów wizyjnych. Scharakteryzowane zostały metody używane do lokalizacji w przestrzeni i tworzenia mapy otoczenia w systemach wizyjnych wykorzystujących obraz stereoskopowy i tych, w których stosowana jest tylko jedna kamera. Pokrewnym problemem jest bezkolizyjna nawigacja autonomicznego robota w nieznanym środowisku. Stosowane są tu metody planowania trasy *bottom-up* polegające na reaktywnej interakcji robota ze środowiskiem i metody *top-down* polegające na szukaniu ogólnego planu trasy, której korekty wprowadzane są w trakcie wykonywania zadania. Wyszukiwanie trasy dla robota często traktowane jest jako zadanie optymalizacyjne, do rozwiązania którego stosowane są algorytmy genetyczne i mrówkowe. Autor omawia też możliwość rozwiązywania problemu nawigacji za pomocą formuł liniowej logiki temporalnej. Następnie Autor rozważa różne metody rozpoznawania obrazów i analizy sceny, kładąc nacisk na metody strukturalne, metody syntaktyczne bazujące na gramatykach ciągowych, drzewowych i grafowych, oraz metody oparte na konwolucyjnych sieciach neuronowych. W dalszej kolejności opisane są wybrane techniki modelowania obiektów w przestrzeni trójwymiarowej. Scharakteryzowane są metody oparte na reprezentacji brył z przesuwaniem i reprezentacjach z podziałem przestrzeni. Przeanalizowane są także techniki segmentacji obrazu umożliwiające identyfikację obiektów w obszarze sceny. W końcowej części rozdziału Autor uzasadnia potrzebę testowania systemów wizyjnych w sztucznym środowisku dla weryfikacji funkcjonowania zarówno modułu kognitywnego jak i zastosowanego protokołu komunikacyjnego.

Rozdział 3 przedstawia strukturę rozprawy i omawia zawartość jej kolejnych rozdziałów.

Rozdział 4 omawia tworzenie reprezentacji wektorowej obiektów znajdujących się na zdjęciu wykonanym przez robota. Obraz wejściowy po preprocessingu ma postać bitmapy, gdzie białe piksele należą do obiektów, a czarne do tła obrazu. W rozdziale przedstawiony jest algorytm wektoryzacji konturów obiektów działający na obrazie wejściowym. Następnie opisany jest zaproponowany przez Autora algorytm wygładzania konturu wektorowego, który zmniejsza zajętość pamięciową reprezentacji. Na zakończenie rozdziału opisany jest postprocessing modelu wektorowego.

Rozdział 5 zawiera opis algorytmu strukturalnego rozpoznawania obrazów, którego pan Śmigieński jest współautorem. Algorytm ten poszukuje zadanego wektorowego modelu obiektu w zbiorze reprezentowanych wektorowo obiektów wejściowej mapy. Wyszukiwanie to działa niezależnie od skali i położenia porównywanych obiektów. Jeśli ilość narożników reprezentacji wektorowej dwóch porównywanych kształtów jest jednakowa, wówczas tworzona jest reprezentacja obiektów bazująca na kątach nachylenia kolejnych fragmentów konturu i ich długości. Jeśli dopasowanie takich reprezentacji zostanie odnalezione następuje jego weryfikacja na podstawie ilorazów długości odpowiadających sobie fragmentów konturów.

Rozdział 6 zawiera opis innego algorytmu strukturalnego rozpoznawania obrazów, którego autorem jest pan Śmigiełski. Algorytm ten nie wymaga założenia, że łamane reprezentujące podobne kształty posiadają taką samą liczbę narożników. Reprezentacje porównywanych konturów tworzone są na bazie pomiarów odległości od środka ciężkości obiektu do punktu konturu dokonywanych co zadany kąt. Ze względu na możliwość różnicy skali porównywanych obiektów dopasowanie reprezentacji ustalane jest na podstawie odchylenia standardowego ilorazów wyników pomiarów.

Rozdział 7 przedstawia metodę konstrukcji trójwymiarowego modelu obiektu na podstawie trzech rzutów bryły wykonanych z różnych kierunków. Pan Śmigiełski jest jednym z autorów tej metody. Na podstawie każdego z trzech rzutów kolejno, generowane są ściany składowe modelu 3D a także dodatkowe cechy obiektu będące otworami na wylot. Na podstawie kolejnych segmentów konturu będącego rzutem odniesienia ustalane są odpowiednie fragmenty kształtów z dwóch pozostałych rzutów, a następnie fragmenty te są rzutowane na płaszczyznę wyznaczoną przez segment odniesienia, prostopadłą do rzutu odniesienia. Część wspólna otrzymanych rzutów tworzy ściany obiektu trójwymiarowego. Zbiór takich ścian w reprezentacji wektorowej stanowi wynik działania opisanego algorytmu. Na końcu rozdziału funkcjonowanie algorytmu zostało zilustrowane na dwóch przykładowych obiektach, dla których wygenerowane zostały modele trójwymiarowe.

W **rozdziale 8** zaprezentowany został opis modułu kognitywnego służącego do tworzenia i analizy modelu sceny. Pan Śmigiełski jest współautorem opisanych tu metod. Rozumienie relacji przestrzennych między obiektami na scenie jest szczególnie istotne w przypadku gdy znajduje się na niej wiele kształtów podobnych do wyszukiwanego. Na podstawie wektorowego obrazu obiektów znajdujących się na wejściowej mapie tworzony jest graf zwany Grafem Bliskiego Sąsiedztwa, którego wierzchołki reprezentują obiekty sceny i przechowują informacje o wektorowym opisie kształtów oraz lokalizacji środków ciężkości obiektów. Krawędzie wychodzące z wierzchołków reprezentują relacje z obiektami znajdującymi się w najbliższym sąsiedztwie, którego zasięg wyznaczany jest za pomocą okręgu o promieniu zaczynającym się w środku ciężkości danego obiektu i kończącym w najbardziej oddalonym punkcie najbliższej położonego obiektu sąsiedniego. Graf taki tworzony jest zarówno dla wyszukiwanego obiektu i jego otoczenia jak i dla obiektów sceny wczytanej przez kamerę robota. Porównanie grafów, w których zapisany jest kontekst przestrzenny obiektów pozwala na znalezienie właściwego obiektu sceny. Obiekty odpowiadające wierzchołkom podgrafu znalezione w grafie reprezentującym scenę są odpowiednio obracane i skalowane. Jeśli pokryją się one z obiektami reprezentowanymi przez poszukiwany podgraf wówczas dopasowanie zostaje znalezione. Jeśli w rozważanym grafie sceny można znaleźć więcej niż jedno takie dopasowanie, kontekst przestrzenny szukanego obiektu może zostać poszerzony o kolejne obiekty sąsiednie.

Rozdział 9 przedstawia opracowany przez Autora moduł nawigacji pozwalający robotowi odpowiednio przemieszczać się w trzech wymiarach. Zaproponowana została metoda obliczania współrzędnych względnych dla punktów docelowych znajdujących się w obszarze sfotografowanym przez robota z dużej wysokości. Punkty te odpowiadają lokalizacjom, znajdującym się bezpośrednio nad każdym z obiektów sceny, do których robot musi się udać w celu wykonania zdjęć tych obiektów. Tak wykonane zdjęcia zapewniają, że rozpoznawane kształty obiektów nie będą posiadały zniekształceń perspektywicznych. Druga zaproponowana tu metoda umożliwi obliczanie docelowych

pozycji znajdujących się z przodu i z boku obiektu, z których robot musi wykonać zdjęcia na potrzeby konstrukcji trójwymiarowego modelu obiektu.

Rozdział 10 zawiera opis sztucznego środowiska testowego, które zostało stworzone przez Autora na potrzeby weryfikacji działania poszczególnych elementów systemu wizyjnego. Podsystem zaimplementowany z wykorzystaniem narzędzia OpenGL umożliwia symulację odbioru danych z kamery robota, wizualizację sceny oraz reakcji symulowanego robota na komendy. Wykorzystanie protokołu komunikacyjnego MavLink do porozumiewania się z jednostką sterującą robota umożliwia bezpośrednio przeniesienie testów z symulatora na prawdziwego robota. Przedstawiony tu scenariusz testowy zakłada wprowadzenie do pamięci robota obiektu, który ma być zlokalizowany na wirtualnej scenie, wykonanie zdjęcia sceny obejmującego wszystkie obiekty, wykonanie zdjęć wszystkich obiektów bezpośrednio z góry, odnalezienie szukanego obiektu a następnie wykonanie jego dwóch dodatkowych zdjęć i stworzenie trójwymiarowego modelu tego obiektu. Opisany został przebieg i wyniki testu oraz przedstawiona droga jaką robot pokonał w wirtualnym środowisku wykonując zadanie.

W **rozdziale 11** omówiona została logiczna i sprzętowa architektura skonstruowanego przez Autora robota latającego oraz testy przeprowadzone z jego udziałem. Testy te zostały wykonane na otwartej przestrzeni z wykorzystaniem makiet budynków pomalowanych na wyróżniające się z otoczenia kolory. Pierwszy z testów polega na wprowadzeniu do pamięci robota kształtu szukanego obiektu wraz z kontekstem przestrzennym i stworzeniu dla niego Grafu Bliskiego Sąsiedztwa, wykonaniu zdjęcia terenu ukazującego rozkład obiektów z wysokości 10 metrów, wyodrębnieniu obiektów o zdefiniowanych wcześniej kolorach, sfotografowaniu tych obiektów bezpośrednio z góry z wysokości 4 metrów, stworzeniu Grafu Bliskiego Sąsiedztwa dla obiektów znajdujących się w terenie, wyszukaniu zadanego obiektu wraz z kontekstem w grafie sceny, wykonaniu dodatkowych zdjęć znalezionej obiektu i stworzeniu trójwymiarowego modelu tego obiektu, oraz sprawdzeniu możliwości przelotu pod obiektem i ewentualnym przemieszczeniu się robota na drugą stronę. W drugim scenariuszu testowym nie jest tworzony trójwymiarowy model znalezionej obiektu, ale spośród obiektów o podobnym kształcie wybierany jest ten właściwy na podstawie jego relacji przestrzennych z innymi obiektami sceny. Przeprowadzone testy weryfikują poszczególne funkcjonalności systemu wizyjnego, które zostały zaimplementowane z wykorzystaniem metod przedstawionych we wcześniejszych rozdziałach rozprawy.

Rozdział 12 zawiera podsumowanie oraz analizuje ograniczenia przedstawionego podejścia i wskazuje kierunki kontynuacji badań mających udoskonalić działanie omawianego systemu wizyjnego.

V. Ocena wyników rozprawy

Trzeba stwierdzić, że cel rozprawy został osiągnięty. Autor zaproponował i przetestował algorytmy dotyczące wektoryzacji obiektów, strukturalnego rozpoznawania kształtów, konstrukcji trójwymiarowego modelu obiektu na podstawie trzech jego rzutów, tworzenia grafowego modelu sceny i analizy sceny poprzez wyszukiwanie podgrafów opisujących relacje przestrzenne między obiektami. Zaproponował środowisko testowe umożliwiające weryfikację działania systemu wizyjnego na symulowanym robocie. Dzięki zastosowaniu protokołu wymiany danych MavLink te same testy zostały wykonane w plenerze na skonstruowanym przez Autora prawdziwym robocie latającym.

Mocną stroną rozprawy jest niewątpliwie podjęcie bardzo ambitnej tematyki dotyczącej analizy i rozumienia sceny przez autonomicznego robota. Autor najwyraźniej zna poruszany temat i jego literaturę oraz ma własne pomysły, które umie należycie opisać i zrealizować. Zaproponowane metody wykorzystane w module kognitywnym robota zostały zweryfikowane praktycznie zarówno na symulowanym robocie w środowisku testowym jak i z wykorzystaniem prawdziwego robota latającego. Ten aspekt rozprawy jest odpowiednio udokumentowany.

Przedstawiony system wizyjny dla robota ma zarówno wartość praktyczną, jak i poznawczą. Rozprawa jest merytorycznie poprawna i spójna logicznie. Język rozprawy jest klarowny i na ogół poprawny.

Rozprawa ma również i słabsze punkty.

1. Rozdział drugi powinien być podzielony na podpunkty, gdyż Autor porusza tu wiele różnych problemów, ale często brak logicznego związku między kolejnymi akapitami. Niezrozumiały jest sposób używania pojęcia 'metody strukturalne i syntaktyczne', jako że syntaktyczne metody rozpoznawania obrazów są podklasą metod strukturalnych. Niepoprawne jest też pojęcie 'automatów parsujących gramatyki'. Dziwi również brak opisu reprezentacji brzegowej i konstruktywnej geometrii brył przy omawianiu technik modelowania złożonych obiektów w przestrzeni trójwymiarowej z wykorzystaniem brył podstawowych.
2. Chociaż w większości przypadków Autor posiada umiejętność opisywania trudnych zagadnień w sposób zrozumiały, to w rozdziale 4 algorytmy opisane są w sposób nieprecyzyjny. W podrozdziale 4.1 okno wykorzystywane do tworzenia konturu wektorowego opisane jest niejasno, wyjaśnienie pojęcia 'punktu początkowego okna' znajduje się zdecydowanie za późno, brak też wyjaśnienia w jakim celu okno jest zamalowywane. Natomiast rysunki mające ilustrować dwa różne przypadki poszukiwania kolejnego punktu konturu są takie same. W podrozdziale 4.2 opisany jest zaproponowany przez Autora algorytm wygładzania konturu wektorowego. Niestety niezrozumiałe jest pojęcie 'odchylenia punktu od prostej'. Jak widać na rys. 4.5 kąty obliczane są pomiędzy wektorami (AB) i (BC) oraz (BC) i (CD), a nie jak twierdzi Autor między wektorem (AB) i punktem C oraz wektorem (BC) i punktem D. Do tymczasowej listy powinny zostać dodane elementy B, q i C, a nie B, q i D. Elementy te dodawane są do listy jeśli suma rozważanych kątów przekroczy zadaną wartość parametru *Angle*. Nie wiadomo jednak czy i co jest dodawane do listy w przeciwnym przypadku, oraz co się stanie gdy suma kątów osiągnie 180° , gdyż w takim przypadku algorytm będzie szukał przecięcia prostych równoległych. Nie jest również jasne co jest przyjmowane jako punkt A w następnej iteracji.
3. Nie wiadomo czemu służy rozdział 5, omawiający jedną z metod strukturalnego rozpoznawania obrazów, gdyż w rozważanym systemie wizyjnym Autor zaimplementował metodę opisaną w rozdziale 6. W przeciwieństwie do metody z rozdziału 5 zapewnia ona efektywne porównywanie i rozpoznawanie kształtów pomimo niedokładności danych wejściowych otrzymanych po preprocessingu.
4. Rysunki 6.4 i 6.5 jako jedyne w pracy opisane są po angielsku, dodatkowo błędnie, gdyż na osi X znajdują się numery pomiarów, a nie wielkość kroku w stopniach. W opisie na rysunku 6.6 brakuje natomiast polskich liter.

5. W podrozdziale 8.2.1 dotyczącym tworzenia Grafu Bliskiego Sąsiedztwa nie została omówiona struktura wierzchołka grafu (rys. 8.3 przedstawia jedynie fragment tej struktury) ani sposób przechowywania jego krawędzi. W kolejnym podrozdziale przedstawiony został problem niepełnego dopasowania, który pojawia się kiedy w grafie sceny zostaje znaleziony podgraf izomorficzny z grafem reprezentującym szukany obiekt wraz z obiektami sąsiednimi, lecz graf bliskiego sąsiedztwa znalezionego obiektu sceny zawiera więcej wierzchołków niż znaleziony podgraf. Nie został natomiast wspomniany problem pokrewny, kiedy niepełne dopasowanie wynika z braku pewnych obiektów na scenie, co jest prawdopodobne gdy przewidywane sąsiedztwo obiektu szukanego może się zmienić w trakcie wykonywania misji przez robota np. w rejonie katastrof.
6. Dla algorytmów omawianych w rozdziałach 5-9 nie została przedstawiona ich złożoność obliczeniowa ani skuteczność działania.
7. Istotnym ograniczeniem przedstawionego podejścia jest konieczność posiadania przez wszystkie obiekty sceny odmiennych kolorów wyróżniających się z otoczenia. Ten warunek umożliwia prawidłowe rozpoznanie obiektów i stworzenie ich trójwymiarowych modeli. Ograniczenie to można byłoby znieść gdyby robot mógł oceniać odległości od obiektów znajdujących się w otoczeniu. Do pozyskiwania wiedzy o otoczeniu robota można wykorzystać czujniki ultradźwiękowe, czujniki podczerwone bądź laserowe skanery dalmierzowe. Często stosowane są systemy stereowizyjne gdzie dzięki zastosowaniu dwóch kamer możliwe jest uzyskanie informacji na temat głębi obrazu, a także pomiar odległości pomiędzy punktami, a co za tym idzie określenie odległości pomiędzy obiektami.

VI. Konkluzja

Podsumowując, oceniam rozprawę pozytywnie i uważam, że świadczy ona o dobrym przygotowaniu Autora do prowadzenia samodzielnej pracy naukowo-badawczej. Pomimo kilku uwag krytycznych praca wnosi istotny wkład w rozwój wiedzy w dyscyplinie naukowej „informatyka” w zakresie systemów wizyjnych dla robotów mobilnych.

Stwierdzam, że recenzowana rozprawa spełnia wymagania, jakie ustawa o stopniach i o tytule naukowym z dnia 14 marca 2003r. przewiduje dla rozpraw doktorskich i na tej podstawie wnoszę o dopuszczenie mgr Piotra Śmigielskiego do publicznej obrony.



