

Politechnika Śląska
Wydział Automatyki Elektroniki i Informatyki
Instytut Informatyki

Recenzja rozprawy doktorskiej

Tytuł rozprawy:

Generowanie modeli widokowych wielościanów monottonnych do identyfikacji wizualnej

Autor rozprawy:

mgr Andrzej Salamończyk

Promotor rozprawy:

Dr hab. inż. Wojciech S. Mokrzycki

Cel zakres i charakter rozprawy

W swym głównym nurcie rozprawa poświęcona jest metodom tworzenia reprezentacji widokowej wielościanów „monotonnych” Warto doprecyzować, że chodzi tu o reprezentację bryły 3D w postaci obrazów 2D lub 2.5D tej bryły lub reprezentację pośrednią wyznaczoną na podstawie tych obrazów. Tworzenie i wykorzystywanie tak rozumianej reprezentacji bryły 3D jest typowe dla człowieka który obiekt 3D stara się zobaczyć ze „wszystkich stron” Autor deklaruje, że utworzone reprezentacje widokowe mogą być podstawą do rozpoznawania typów/rodzajów brył na podstawie ich widoków co znowu jest próbą naśladowania funkcjonalności człowieka który po obejrzeniu bryły 3D i zapamiętaniu wyników oglądu będzie w stanie w przyszłości rozpoznać ten typ bryły. Dodatkowym nurtem rozprawy jest wartościowy przegląd literatury w zakresie reprezentacji brył nakierowany na ich identyfikację wizualną znacznie wykraczający zarówno zakresem jak i precyzją poza standard rozpraw doktorskich. W szerszym ujęciu tematykę rozprawy można sklasyfikować jako rekonstrukcję semantyki sceny 3D na podstawie obrazów 2D lub 2.5D.

Podsumowując oceniam, że tematyka rozprawy wybrana została ambitnie i stanowi rodzaj wyzwania intelektualnego w obszarze geometrii. Ma również znaczenie praktyczne dla wizji komputerowej (rozumienie sceny) i robotyki. W ramach rozprawy dokonano poprawnego i wyczerpującego przeglądu istniejących metod tworzenia reprezentacji brył przeznaczonych do ich klasyfikacji.

Zawartość rozprawy i jej dyskusja

Rozprawa liczy 92 strony, napisana została w języku polskim, składa się z 5 rozdziałów, wykazu literatury, spisów: rysunków, oznaczeń, skrótów, oraz dwóch dodatków. Niewielka objętość rozprawy jest konsekwencją założonej dyscypliny formalnej. W rozprawie nie ma zbędnych zdań i opisów. Rozprawa posiada bardzo dobry, jasny i intuicyjny liniowy układ treści. W układzie treści rozprawy można wyróżnić dwie części. Pierwsza zawiera materiał

wprowadzający do części badawczej. Część druga zawiera opis i badania wykonane w zakresie reprezentacji widokowej.

W rozdziale 1 omówiono koncepcję rozprawy i rolę reprezentacji widokowej w Systemach Identyfikacji Wizualanej. Ten fragment rozprawy można widzieć jako wprowadzenie do rozdziału 4 i tym samym wyjaśnienie po co tworzona i jak może być wykorzystana reprezentacja widokowa. Autor drobiazgowo omawia proces tworzenia baz danych widoków i ich typy, szkoda, że pomija choćby skrótowe omówienie samej koncepcji pasowania. Czytelnika który już polubił skondensowany i precyzyjny język narracji mogą razić odstępstwa od tej konwencji przykładowo w postaci „Rozpoznawanie to proces pasowania pomiędzy wyglądem sceny i opisem modelu obiektu” Co to jest wygląd sceny? dlaczego precyzyjne pojęcie bryła 3D (element sceny) zastąpiono pojęciem sceny czyli fragmentem przestrzeni 3D a reprezentację „opisem” Istotnym fragmentem omawianego rozdziału jest punkt 1.2 zatytułowany „Podstawowe pojęcia i definicje” Autor realizując przyjętą konwencję rozprawy postanowił zdefiniować używane dalej terminy. Podoba mi się wykazany umiar polegający na rezygnacji z manieri tworzenia dziesiątków definicji w rodzaju Definicja 1,..., a stosowaniu wytluszczenia definiowanych terminów w litym tekście. Doceniam również wykazaną w ten sposób troskę Autora o czytelnika w sytuacji bardzo specyficznego i charakterystycznego podejścia do terminologii przejętego zapewne od Promotora.

Podano podstawową definicję sfery widokowej w przypadku rzutowania perspektywicznego i związane z nią definicje punktu widokowego, stożka widokowego, i obszaru jednowidokowego. Dla potrzeb wspomnianych definicji z bryłą wiązany jest lokalny kartezjański układ współrzędnych/ramka z początkiem w środku geometrycznym bryły. W mojej ocenie należało jeszcze podać definicję środka geometrycznego bryły. Jeżeli bryła posiada symetrię środkową nie ma problemu z wyznaczeniem środka geometrycznego ale jak wyznaczyć ten środek przykładowo dla bryły ze str. 91? Ponieważ wybór lokalnej ramki bryły definiuje małą sferę jak również pośrednio sferę widokową powstaje pytanie jaki jest wpływ wyboru początku tej ramki na cały proces wyznaczania widoków. Istotną dla koncepcji rozprawy jest definicja widoku. Dlaczego Autor używa terminu „płaty powierzchniowe” skoro ograniczył klasę brył do wielościanów „monotonnych”? Warto było określić jak reprezentowane są/będą „płaty powierzchniowe” lub odwołać się do przykładów i list z dodatku B. W punkcie 1.3 podano podstawowe założenia przy których rozpatrywany jest cały problem oraz tezy rozprawy. W fragmencie dotyczącym założeń Autor umieścił zaczerpnięte z literatury funkcje określające liczby widoków w zależności od liczby ścian oraz złożoności obliczeniowej podziału przestrzeni widokowej i wyznaczenia aspekt-grafu. W obydwu przypadkach funkcje oznaczane są tą samą literą „duże O” co w złożoności obliczeniowej ma swoje precyzyjne znaczenie natomiast w przypadku liczby widoków należało to znaczenie określić. Proszę o wyjaśnienie tego podczas obrony i wzbogacenie o definicje złożoności obliczeniowej w tym „małe o”. Tezy rozprawy sformułowane w zakończeniu punktu 1.3 są poprawne w tym znaczeniu, że może być podany ich dowód drobną usterką jest to, że zawierają terminy które zostaną wyjaśnione dopiero w dalszej części rozprawy.

Rozdział 2 zatytułowany „Przegląd literatury” zawiera wartościowe wyniki w postaci wyczerpujących opisów i precyzyjnych analiz w zakresie reprezentacji bryły wytworzonych na podstawie jej obrazu/obrazów 2D, 2,5D. Podobnie jak w rozdziale 1 Autor na początku rozdziału 2 również stara się zdefiniować lub doprecyzować terminy których będzie używał w dalszej części. Jako wartościowe oceniam też odwołania do poglądów zaczerpniętych ze znaczących publikacji. W treści tego rozdziału wyróżnić można dwa wątki. Pierwszy to przegląd ogólny reprezentacji/opisu brył dla potrzeb ich rozpoznawania, drugi to historia i

podsumowanie podejścia szczególnego opartego na podziale sfery widokowej na obszary jednowidokowe zapoczątkowanego przez Promotora w latach 1995 [23, 24, 25, 26] i konsekwentnie kontynuowanego aż do chwili bieżącej. Jeśli chodzi o ogólny przegląd reprezentacji i opisów brył dla potrzeb ich identyfikacji wizualnej to ze względu na jego dojrzałość, zastosowaną systematykę podejść widzę możliwość redakcji tego fragmentu do postaci książkowej po rozszerzeniu opisów i dodaniu przykładów. Szczególnie ciekawy i wartościowy jest podpunkt 2.2.4 poświęcony deskryptorom kształtu. Jak słusznie zauważa Autor termin ten wraz z różnymi technikami wyznaczania deskryptorów (np. SIFT, SURF) jest znany i używany w wizji komputerowej. Przedstawione opisy tworzenia deskryptorów brył 3D stanowią rozszerzenie deskryptorów 2D i zasługują jak już pisałem wcześniej na popularyzację. Zaletą przeglądu zamieszczonego w p.2.3 są drobiazgowo prezentacje koncepcji i algorytmów niekiedy bardzo skomplikowanych związanych z koncepcją widoków oraz propozycje udoskonaleń lub implementacji. Przegląd rozpoczęto od grafów aspektowych które w mojej ocenie stanowią najlepszą reprezentację topologii bryły o czym świadczy również cytowana przez Autora literatura. Grafy aspektowe są „elastyczne” i pozwalają na pominięcie w pierwszej fazie rozpoznawania atrybutów ilościowych takich jak skala, kąty, powierzchnie. W drugiej fazie można używać widoków. Dla spójności z zawartością rozdziału 4 warto również podkreślić, że już w wczesnych pracach pod kierunkiem Promotora pojawiało się pojęcie widoków 2.5 D wraz z koncepcją ich użycia w rozpoznawaniu brył. Autor doprecyzowuje pojęcie widoku dodając do widocznej w widoku listy ścian wielościanu, współrzędne (w ramce globalnej) wierzchołków wielokątów stanowiących te ściany. W dalszym ciągu w p.2.4 omawiane są algorytmy generowania reprezentacji widokowej z rzutowaniem środkowym [23] w wersji iteracyjnej z zakończeniem iteracji w przypadku pokrycia całej sfery widokowej obszarami jednowidokowymi. Kolejnym etapem rozwoju reprezentacji bryły przez widoki były algorytmy nie iteracyjne oparte na koncepcji rozrostu pojedynczego obszaru jednowidokowego. Rozrost obszarów jednowidokowych kończy się po stwierdzeniu, że każdy obszar jednowidokowy ma sąsiada wzdłuż każdej swojej krawędzi. Wyznaczanie pojedynczego obszaru jednowidokowego, ziarna rozrostu, realizowane jest przez śledzenie jego granic do momentu powrotu do punktu startowego. W kolejnych pracach algorytm śledzenia granic został rozszerzony na przypadek wielościanów „monotonnych” Prezentacja stosowanych wcześniej koncepcji i realizujących je algorytmów stanowiła przygotowanie do algorytmu zaproponowanego przez Autora. Algorytm ten jest udoskonaleniem algorytmu opartego na wektorowej reprezentacji wielościanu i koncepcji stożka dopełniającego. Stożek dopełniający w nowej wersji obraca się względem kolejnych krawędzi ostrosłupa danej ściany o kąty zależne od katów płaskich w wierzchołkach tej ściany jest oraz długości krawędzi (ściany widokowej) stanowiących ramiona tych kątów.

Rozdział 3 zawiera podstawowe wyniki rozprawy. Jego tytuł „Przybliżone metody generowania reprezentacji widokowej” jest dyskusyjny. Autor w zakończeniu poprzedniego rozdziału 2 pisze „Sposoby generowania reprezentacji widokowej możemy podzielić na dokładne (generujące kompletny zestaw widoków) i przybliżone w których widoki są wyznaczane z dokładnością do określonej rozdzielczości” Ponieważ w rozprawie nie sprecyzowano jaką techniką pozyskiwane są widoki a w rozdziale 4 mówi się o mapach głębi można przypuszczać, że chodzi o wielkość progową ściany wielościanu a nie rozdzielczość pikselową ekranu. W krótkim pierwszym akapicie rozdziału 3 Autor rezygnuje z wyznaczania obszarów jednowidokowych i skupia się jedynie na pozyskiwaniu widoków bryły. Ogólna koncepcja wszystkich algorytmów prezentowanych w tym rozdziale to próbkowanie sfery widokowej według różnych strategii (punkty krzywej parametrycznej leżącej na powierzchni sfery, wierzchołki wielościanu foremnego tworzone techniką podpodziału również w wersji wielorozdzielczej) wyznaczanie dla danego punktu widoku i porównywanie z widokami ze

zbioru widoków już istniejących i dołączanie do zbioru widoków jeśli nie ma w nim takiego widoku. W ocenie złożoności obliczeniowej w p.3.5 Autor wspomina o sortowaniu widoków co niewątpliwie uprościłoby fazę przeszukiwania zbioru widoków. Proszę o wyjaśnienie jak takie sortowanie jest realizowane? Algorytmy opisane w rozdziale 4 są poprawne i zapewne praktyczne jednak ich „złożoność logiczna” jest znacznie niższa niż algorytmów omawianych w rozdziale 3 a w szczególności ostatniego z nich.

W rozdziale 4 Autor pokazuje jak mogą być wykorzystane zbiory widoków bryły do jej identyfikacji. W pełni zgadzam się z klasyfikacją podaną w podpunkcie 4.1.1. Ostatecznie do dalszych rozważań przyjmowana jest reprezentacja widoku w której do ścian przypisane są ich cechy geometryczne (powierzchnia, długości boków, kąty pomiędzy bokami, normalna) Niepokój budzi fakt, że w rozprawie nie ma fragmentów ani nawet komentarza jak są pozyskiwane cechy geometryczne. Inną sprawą jest pytanie dlaczego właśnie takie składowe reprezentacji wektorowej zostały wybrane. Lepszą spójność z wcześniejszymi fragmentami rozprawy daje zbiór widoków z indeksami ścian. Oczywiście nadal pozostaje potrzeba określenia jak rozróżniane są ściany. Ogólną usterką rozprawy jest to, że nie zostały wcześniej omówione techniki pozyskiwania widoków. W mojej ocenie preferowaną jest tu technika obrazów 2D z dobrze ugruntowanymi algorytmami wyznaczania krawędzi i segmentacji. Autor w nienumerowanym podpunkcie „Pozyskiwanie danych” proponuje technikę map głębokości. Popełnia jednocześnie być może nieświadomie błąd pisząc „można otrzymać odległości od uporządkowanego zbioru punktów z otoczenia skanera” Pomimo, że wszystkie techniki opisane w rozdziale 4 oceniam jako poprawne i świadczące o rozeznaniu i erudycji Autora to moim zdaniem koncepcja przedstawiona w tym rozdziale nie tworzy racjonalnej całości z wcześniejszymi fragmentami pracy. Przykładowo posiadając modele/modelę brył w postaci siatki powierzchni a nie zbioru widoków pasowanie z chmurą punktów ze skanera wyglądałoby identycznie jak to opisano w rozprawie co więcej łatwo byłoby wprowadzić dodatkowe stopnie swobody takie jak skalowanie ścinanie jak również pasowanie hierarchiczne wykorzystujące siatkę utworzoną na chmurze punktów i koncepcję wielorozdzielczości.

Ostatnim rozdziałem rozprawy jest „Podsumowanie” w którym w zwięzły sposób przedstawiono tło rozprawy jak i oryginalne wyniki.

Oryginalne rezultaty rozprawy

Oryginalnymi wynikami w zakresie koncepcji są: i)strategia ruchu stożka dopełniającego polegająca na jego obrocie wokół krawędzi ostrosłupa ściany, ii)próbkowanie punktów widokowych wzdłuż krzywej spiralnej leżącej na powierzchni sfery widokowej, iii)próbkowanie punktów widokowych z wykorzystaniem wielościanów foremnych, iv)pasowanie reprezentacji widokowej z mapami zasięgu. Sformułowane koncepcje były podstawą do opracowania czterech nowych algorytmów generowania reprezentacji widokowej bryły 3D. Algorytmy zostały zaimplementowane i przebadane numerycznie. Wynikami teoretycznymi a zarazem potwierdzeniem słuszności postawionych tez są oszacowania złożoności obliczeniowej algorytmu tworzenia reprezentacji widokowej i algorytmu pasowania reprezentacji widokowej z mapą/mapami głębokości bryły.

Analiza źródeł

Rozprawę cechuje bardzo obszerna, celna i aktualna bibliografia świadcząca o doskonałym rozeznaniu literaturowym wykazanym między innymi w rozdziale „Przegląd literatury” będącym w rzeczywistości samodzielny studium badawczym. Podobną biegłość w korzystaniu z literatury odnotowałem również w całej rozprawie. W konkluzji oceniam, że

analiza źródeł i ogólna umiejętność korzystania z literatury wykazana przez Autora jest bardzo dobra z wyróżnieniem.

Znaczenie uzyskanych wyników

Wyniki uzyskane w rozprawie mają znaczenie zarówno teoretyczne jak i praktyczne ale w mojej ocenie są głównie potwierdzeniem wysokich możliwości intelektualnych doktoranta.

Redakcja rozprawy

Rozprawę cechuje bardzo dobra logika i redakcja. Szczególnie wysoko oceniam przyjęty formalny styl rozprawy okupiony niekiedy drobnymi potknięciami zrozumiałymi jeśli próbuje się zbudować zamknięty system logiczny. Sporadycznymi błędami jest brak konsekwencji terminologicznej oraz drobne błędy edycyjne. Edycyjną zaletą rozprawy jest użycie profesjonalnego edytora tekstu.

Omawiając zawartość rozprawy nie komentowałem specyficznej terminologii wprowadzonej przez Autora zapewne pod wpływem Promotora. W rozprawie jest ona spójna i oceniam, że Autor ma prawo do wprowadzania własnych terminów.

Uwagi i słabe strony rozprawy

Rozprawa nie posiada błędów merytorycznych. Drobne błędy logiczne, językowe i literowe naniósłem w przekazanym mi egzemplarzu rozprawy. Jako temat do dyskusji, oprócz wymienionych w omówieniu zawartości rozprawy, proponuje omówienie pasowania w wersji gdy zamiast widoków 2.5D dostępne są obrazy 2D.

Podsumowanie

Podsumowując całość rozprawy stwierdzam, że Autor wykazał się doskonałym opanowaniem warsztatu naukowego, pracowitością i rzetelnością naukową oraz wysokim poziomem intelektualnym który pozwolił na opanowanie historii i posunięcie o krok do przodu rozwiązania „łamiągłówki” geometrycznej wyznaczenia kompletnego zbioru widoków wielościanów monottonnych.

Ocena końcowa

Stwierdzam, że w recenzowanej rozprawie, został poprawnie sformułowany, a następnie rozwiązany z zastosowaniem metod naukowych, złożony problem wyznaczenia widoków wielościanów monottonnych w kilku alternatywnych podejściach. Oceniana rozprawa doktorska spełnia wymagania, jakie Ustawa o Stopniach i o Tytule Naukowym przewiduje dla rozpraw doktorskich. Wobec powyższego wnioskuję o jej przyjęcie jako rozprawy doktorskiej i dopuszczenie do dalszych etapów przewodu doktorskiego.

K. Węgielowski