dr hab. Przemysław Kiciak
Instytut Matematyki Stosowanej i Mechaniki,
Uniwersytet Warszawski,
ul. Banacha 2
02-097 Warszawa

Warszawa, 9 lutego 2012

Recenzja rozprawy doktorskiej

dla Rady Naukowej
Instytutu Podstawowych Problemów Techniki
Polskiej Akademii Nauk

Generowanie modeli widokowych wielościanów monotonnych do identyfikacji wizualnej

mgra Andrzeja Salamończyka

1. Jakie zagadnienie naukowe jest rozpatrzone w pracy (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Tematem rozprawy jest tworzenie reprezentacji brył wielościennych, umożliwiającej rozpoznawanie takich brył na podstawie obrazów zarejestrowanych przez kamerę. Po dokonaniu segmentacji obrazu i wyodrębnieniu w nim odcinków i wielokątów, będących rzutami perspektywicznymi widocznych fragmentów krawędzi i ścian bryły, otrzymuje się tzw. widok wielościanu, który zawiera informację o topologii i geometrii fragmentu brzegu wielościanu przedstawionego na obrazie. Informacja ta może być następnie dopasowana do modeli widokowych przechowywanych w bazie danych, w celu rozpoznania obiektu na podstawie jego obrazu. Najważniejszą częścią pracy jest opis proponowanego przez autora algorytmu tworzenia modeli widokowych wielościanów, z których można utworzyć bazę danych umożliwiającą rozpoznawanie tych wielościanów.

Motywacją do rozwiązania tego problemu jest zastosowanie w robotyce; robot przemysłowy wyposażony w system wizyjny, dzięki zdolności do identyfikacji przedmiotów, może skuteczniej nimi manipulować np. na stanowisku montażowym. W tym celu baza danych tego systemu powinna zawierać modele widokowe przedmiotów, którymi ma manipulować robot.

Możliwość rozpoznania i złożoność obliczeniowa tego zadania w istotny sposób zależą od cech kształtu bryły. Z uwagi na cechy istotne dla rozpoznawania bryły wielościenne można podzielić na różne klasy, np. klasę wielościanów wypukłych, klasę wielościanów gwiaździstych, klasę graniastosłupów i inne. Algorytmy proponowane w pracy mają zastosowanie do klasy wielościanów monotonnych. Klasa ta składa się z brył gwiaździstych ze względu na wyróżniony punkt i w szczególności zawiera klasę wielościanów wypukłych.

Praca ma charakter badawczo-implementacyjny. Opracowanie algorytmu tworzenia modeli widokowych wielościanów należących do ustalonej klasy wymaga dokonania teoretycznej analizy problemu. Analiza taka jest zamieszczona w rozprawie, a jej wyniki są podstawą do przyjęcia konkretnych sposobów rozwiązywania poszczególnych podproblemów w zaproponowanym algorytmie, który został zrealizowany i przetestowany na przykładach.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle) świadczący o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonywujący?

Rozdział 2 pracy zawiera przegląd literatury związanej z treścią pracy. Znanych jest wiele metod reprezentowania brył. Metody te zostały opracowane na potrzeby rozmaitych zastosowań, m.in. grafiki komputerowej, której celem jest synteza obrazów, projektowania wspomaganego komputerem (CAD/CAM), gdzie modele obiektów są podstawą do ich wytwarzania za pomocą obrabiarek sterowanych numerycznie, modelowania matematycznego, gdzie bryła jest obszarem, w którym mają miejsce pewne zjawiska fizyczne (np. naprężenia albo przepływy ciepła lub płynów) opisane przez równania różniczkowe, które należy rozwiązywać numerycznie, i wreszcie przetwarzania i rozpoznawania obrazów, w którym należy zidentyfikować obiekt na podstawie jego cyfrowej fotografii.

Każda z wymienionych wyżej dziedzin ma swoją specyfikę, wskutek czego stosowane w nich modele obiektów mają rozmaite własności, od których zależy łatwość albo trudność użycia w konkretnych zastosowaniach. Poszczególne sposoby reprezentowania obiektów są odpowiednie dla różnych klas obiektów, zawierają rozmaite informacje podstawowe i dodatkowe (w tym redundantne, istotne dla złożoności obliczeniowej algorytmów) i są w naturalny sposób związane ze metodami uzyskiwania danych opisujących konkretny obiekt. Dane te mogą stanowić w skrajnym przypadku zbiór niepowiązanych punktów na powierzchni bryły, otrzymanych za pomocą skanera laserowego. Inny przykład reprezentacji to tzw. drzewo CSG (ang. *Constructive Solid Geometry* — konstrukcyjna geometria brył), które opisuje bryłę powstającą z obiektów pierwotnych reprezentowanych przez liście, przy czym wierzchołki wewnętrzne drzewa reprezentują regularyzowane operacje mnogościowe (wynikiem takiej operacji jest domknięcie wnętrza sumy, różnicy, przecięcia, lub dopełnienia argumentów reprezentowanych przez poddrzewa). Analizowanych w pracy metod reprezentacji jest znacznie więcej niż wymienione wyżej. Pewne metody dopuszczają tylko reprezentowanie wielościanów o ścianach płaskich, inne także obiektów o powierzchniach zakrzywionych (za pomocą płatów parametrycznych lub w postaci niejawnej). Poszczególne reprezentacje mogą zawierać dodatkowe informacje o obiekcie, na przykład o jego topologii, symetriach itp.

Przegląd sposobów reprezentowania brył dokonany w recenzowanej pracy jest bardzo obszerny, ale w całym rozdziale 2 uwaga pozostaje skupiona na cechach modeli istotnych dla rozwiązywanego zadania, tj. rozpoznawania obiektów na obrazach. Uzasadnienie wybrania modeli widokowych dla tego zastosowania jest jasne: na obrazie bryły wielościennej można, za pomocą segmentacji, wyróżnić odcinki będące obrazami krawędzi i wielokąty będące obrazami widocznych fragmentów płaskich ścian. Sąsiedztwo ścian mających wspólne krawędzie jest podstawową informacją przechowywaną w modelu widokowym, a zatem informacja otrzymana w wyniku segmentacji obrazu może być bezpośrednio dopasowywana do informacji w modelach brył, które należy rozpoznawać na obrazie.

W dalszej części rozdziału 2 rozprawy autor dokonał przeglądu i klasyfikacji metod generowania widoków brył wielościennych na potrzeby identyfikacji wizualnej. Klasyfikacja ta jest oparta na dwóch podziałach zbioru znanych metod: wyróżnia się metody znajdujące obszary jednowidokowe i metody, które tego nie robią. W obu tych klasach wyróżnia się tzw. metody iteracyjne, które kończą działanie po spełnieniu określonych warunków wskazujących osiągnięcie celu, i metody zwane nieiteracyjnymi, które wykonują z góry zadaną liczbę kroków obliczeniowych, np. znajdują widoki bryły z punktów widokowych, których zbiór jest ustalony zawczasu. Ta część przeglądu rozwiązań znanych z literatury jest bezpośrednio związana z tematem pracy i zawiera bardziej szczegółowe opisy algorytmów. Niektóre z tych algorytmów są oryginalnym dorobkiem autora rozprawy, opisanym w jego wcześniejszych publikacjach.

3. Czy autor rozwiązał postawione zagadnienie, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Istnieją dwa ogólne podejścia do rozwiązywania praktycznych zadań geometrii obliczeniowej. Algorytmy rozważane w pierwszym z tych podejść są teoretycznie dokładne, tj. wynik obliczenia jest pełną informacją o rozwiązaniu postawionego zadania. W zadaniu rozpatrywanym w recenzowanej pracy takim wynikiem byłby podział tzw. sfery widokowej, wewnątrz której jest umieszczony obiekt i w której dowolnym punkcie można umieścić obserwatora (czyli środek optyczny obiektywu kamery, która ma za zadanie sfotografować obiekt), na obszary jednowidokowe.

Obszar jednowidokowy jest zbiorem punktów, z których na powierzchni obiektu „widać to samo", tj. zbiory wierzchołków, krawędzi i ścian widocznych w całości oraz krawędzi i ścian widocznych częściowo są identyczne. Dla każdej ściany bryły monotonnej istnieją kierunki, z których widoczna jest cała ta ściana i co najmniej fragmenty ścian sąsiednich. Między innymi dzięki temu znalezienie widoku bryły z dowolnego punktu (reprezentanta) każdego obszaru jednowidokowego daje gwarancję możliwości dopasowania informacji otrzymanej z dowolnego obrazu tej bryły utworzonego przez kamerę umieszczoną w dowolnym punkcie sfery widokowej do któregoś z tych widoków.

Algorytmy dokładne, w tym algorytmy rozwiązujące postawione w pracy zadanie w sensie opisanym wyżej, często mają istotne wady utrudniające ich praktyczne stosowanie. Liczba obszarów jednowidokowych nawet dla wielościanów o stosunkowo małej liczbie ścian jest bardzo duża, ponadto dla brył niewypukłych (w tym rozważanych w pracy wielościanów monotonnych) niektóre krzywe rozgraniczające obszary jednowidokowe nie są łukami okręgów na sferze widokowej (jest to związane z opisanymi w pracy tzw. zdarzeniami widokowymi typu *kkk*), co komplikuje reprezentację wyniku obliczeń. Algorytmy dokładne mają zwykle zbyt dużą dla zastosowań praktycznych złożoność obliczeniową. Ponadto nawet gdyby takie dokładne rozwiązanie zadania było dostępne, fakt, że identyfikacja obiektu ma być przeprowadzona na podstawie obrazu rastrowego, który z natury rzeczy nie zawiera dokładnej informacji o widocznych obiektach, w istocie likwiduje wspomnianą wyżej gwarancję poprawnego dopasowania obrazu do widoku.

Z powodów przedstawionych wyżej znacznie większe znaczenie praktyczne mają algorytmy aproksymacyjne z elementami heurystyki. Rozwiązanie proponowane w pracy polega na wybraniu a priori pewnego zbioru punktów na sferze widokowej i wygenerowaniu (na podstawie modelu w postaci reprezentacji brzegowej) widoków bryły z tych punktów. Uzyskany w ten sposób zbiór widoków może zawierać powtórzenia (które są eliminowane za pomocą sortowania), przy czym, zależnie od złożoności kształtu, istnieje dostatecznie gęsty skończony zbiór punktów widokowych, dla którego otrzymuje się komplet widoków bryły.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy, czy poziomu techniki reprezentowanych przez literaturę światową?

Zaproponowana w pracy metoda wybierania zbioru punktów widokowych polega na skonstruowaniu krzywej spiralnej leżącej na sferze widokowej; końce tej krzywej są punktami antypodycznymi sfery. Punkty widokowe są rozmieszczone na tej krzywej i dzielą ją na łuki o równych długościach.

Znana jest z literatury metoda podziału trójkątnych ścian dwudziestościanu foremnego. Każdą ścianę dwudziestościanu można podzielić na n^2 trójkątów, przy czym liczba n musi być naturalna. Liczba otrzymanych w ten sposób punktów, odpowiadających wierzchołkom trójkątów i dzięki temu rozmieszczonych równomiernie na sferze widokowej, jest równa $12 + 30(n - 1) + 10(n - 2)(n - 3)$. Nie każda liczba naturalna może być wartością tego wyrażenia.

W metodzie proponowanej w rozprawie liczba punktów, które również są rozmieszczone równomiernie na sferze widokowej, jest sterowana parametrem σ, określającym gęstość zbioru punktów widokowych. Parametr ten wyznacza m.in. kształt krzywej spiralnej i może przyjmować dowolne rzeczywiste wartości dodatnie. Opisana wzorem (3.8) w pracy zależność liczby otrzymanych tą metodą punktów widokowych od parametru σ umożliwia dobranie jego wartości prowadzące do otrzymania dowolnej założonej liczby punktów. Tak więc proponowana w pracy metoda jest bardziej elastyczna niż metody podziału ścian wielościanów foremnych.

Rozdział 4 rozprawy zawiera przegląd metod dopasowania informacji otrzymanej z obrazu do widoków obiektów, które należy rozpoznać. Jedna z tych metod, opisana bardziej szczegółowo, została opracowana przez autora rozprawy i opisana w jego wcześniejszej publikacji. Szczególna uwaga w opisie tej metody jest poświęcona możliwości jej stosowania do rozpoznawania wielościanów monotonnych. W ten sposób rozprawa stanowi zamkniętą całość. Środkiem do rozwiązania postawionego problemu rozpoznawania obiektów należących do wybranej klasy (wielościanów monotonnych) na podstawie obrazów jest informacja (zgromadzona w bazie danych z widokami), otrzymana metodą przedstawioną w rozdziale 3. Rozdział 4 zawiera opis metod wykorzystania tej informacji do rozpoznawania obiektów z tej klasy.

5. Czy autor wykazał umiejętność poprawnego i przekonywującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

6. Jakie są słabe strony rozprawy i jej główne wady?

Układ pracy, tj. kolejność przedstawianych w niej problemów i opis sposobu ich rozwiązywania, a także uzasadnienie sformułowanych w rozprawie tez i wniosków oceniam jako poprawne. Znalazłem w pracy trzy usterki.

W rachunkach na stronach 52–53 jest luka. Autor użył niestandardowej definicji całki eliptycznej drugiego rodzaju i nie podał tej definicji, a ponadto pominął część rachunków w wyprowadzeniach przedstawionych w pracy wzorów. Najczęściej używana definicja zupełnej całki eliptycznej drugiego rodzaju jest następująca: jest to funkcja

$$E(p) = \int_0^1 \sqrt{\frac{1 - p^2x^2}{1 - x^2}}\, \mathrm{d}x = \int_0^{\pi/2} \sqrt{1 - p^2 \cos^2 t}\, \mathrm{d}t \quad \text{dla } p \in [0, 1].$$

Na podstawie parametryzacji (3.6) krzywej spiralnej można otrzymać zależność długości L tej krzywej od parametru k; jeśli promień sfery $R = 1$, to

$$L(k) = \int_0^{\pi} \sqrt{1 + k^2 \sin^2 t}\, \mathrm{d}t = 2 \int_0^{\pi/2} \sqrt{k^2 + 1 - k^2 \cos^2 t}\, \mathrm{d}t.$$

Przekształcenie tego wzoru prowadzi do otrzymania formuły

$$L(k) = 2\sqrt{k^2 + 1} E\big(k/\sqrt{k^2 + 1}\big),$$

w której występuje całka eliptyczna zdefiniowana wzorem podanym wyżej, natomiast podane przez autora wyrażenie ma postać $L = 2E(-k^2)$, przy czym $k \gg 1$. Niemniej, oszacowanie długości krzywej spiralnej (3.12), będące końcowym wnioskiem z zamieszczonego w pracy rachunku, jest poprawne.

Dowód twierdzenia 3.3.1 nie jest poprawny, choć jest oparty na poprawnym pomyśle. Jeśli łuki okręgów wielkich łączące końce krzywej spiralnej nazwiemy południkami, to kąt przecięcia krzywej spiralnej z południkiem w żadnym punkcie nie jest kątem prostym, choć poza otoczeniem punktów końcowych krzywej dla $\sigma \to 0$ miara tego kąta dąży do $\pi/2$. Punkt P_3, będący przecięciem południka, na którym leży punkt widokowy P_1, i kolejnego zwoju krzywej spiralnej, na ogół nie jest punktem widokowym. Dlatego wykazanie, że co najmniej jeden z punktów, P_1, P_2 lub P_3, leży wewnątrz stożka o kącie rozwarcia 2σ, na którego osi leży punkt P_0, *nie dowodzi*, że stożek ten zawiera co najmniej jeden punkt widokowy — wystarczyłoby jednak udowodnić, że jeśli punkt krzywej spiralnej położony najbliżej punktu P_0 leży na łuku (tej krzywej), którego końcami są punkty widokowe P_1 i P_2, to rozpatrywany stożek zawiera co najmniej jeden z tych punktów.

Trzecia usterka to zbyt daleko idący wniosek w podsumowaniu na stronie 70. Autor napisał: „Przedstawione w pracy metody przybliżone generowania reprezentacji widokowej są na tyle uniwersalne, że po niewielkiej modyfikacji (...) mogą być z powodzeniem stosowane dla brył innych rodzajów. Szczególnie dotyczy to ogólnej klasy wielościanów, ale można je również wykorzystywać dla innych obiektów.” Można się domyślić, że ostatnie zdanie dotyczy brył, których brzeg składa się z powierzchni zakrzywionych. Widoczna część brzegu takiej bryły jest ograniczona nie tylko przez krawędzie (których bryła może w ogóle nie mieć), ale także przez krzywe konturowe, które dla każdego punktu widokowego mogą być inne. Dla takich brył należałoby odpowiednio rozszerzyć pojęcie zdarzenia widokowego i opracować algorytmy dostosowane do tak uogólnionego zadania, co jest nietrywialnym problemem otwartym. Zatem to stwierdzenie nie jest dostatecznie uzasadnione.

Z uwagi na implementacyjny charakter pracy oraz wiele zawartych w niej wartościowych wyników nie uważam opisanych wyżej usterek za istotne.

7. Jaka jest przydatność rozprawy dla nauk technicznych?

System identyfikacji wizualnej zrealizowany z wykorzystaniem algorytmów proponowanych w rozprawie może być użyty do sterowania robotami przemysłowymi pracującymi w wielu różnych zastosowaniach. Klasa rozpoznawanych obiektów — brył monotonnych — jest istotnie szersza od klasy wielościanów wypukłych i jest dostatecznie szeroka, aby taki system identyfikacji wizualnej miał duże możliwości przydatne w praktyce. Dlatego moja ocena przydatności rozprawy w naukach technicznych jest wysoka.

8. Do której z następujących kategorii Recenzent zalicza rozprawę?

a) niespełniająca wymagań stawianych rozprawom doktorskim przez obowiązujące przepisy

b) wymagająca wprowadzenia poprawek i ponownego recenzowania

c) spełniająca wymagania

d) spełniająca wymagania z wyraźnym nadmiarem

e) wybitnie dobra, zasługująca na wyróżnienie

Podsumowując, stwierdzam, że praca spełnia wymagania stawiane rozprawom doktorskim i wnoszę o dopuszczenie jej do obrony.