

RECENZJA ROZPRAWY DOKTORSKIEJ

Keypoint matching for object detection in 2.5D images

(Dopasowanie punktów charakterystycznych w obrazach 2.5D do wykrywania obiektów)

Autor rozprawy: mgr inż. Karol Matusiak

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)?

Tematyka rozprawy doktorskiej mgr inż. Karola Matusiaka zawiera się w dziedzinie rozpoznawania obrazów. W dalekiej perspektywie rozpoznawanie obrazów jest częścią obszaru sztucznej inteligencji, konkretnie - sztucznego zmysłu wzroku, niemniej już dzisiaj ma ono znaczenie w robotyce, wojskowości, medycynie, analizie baz zdjęć itp. Rozpoznawanie obiektów w obrazie rozpoczyna się od znajdowania i dopasowywania związanych z nimi punktów charakterystycznych. Efektywność metod rozpoznawania i dopasowywania tych punktów znacznie ustępuje umiejętnościom człowieka, tak więc jest to ciągle ważny temat badań naukowych. Istotny postęp nastąpił w dziedzinach wycinkowych, jak na przykład w rozpoznawaniu twarzy, przedstawione w pracy algorytmy dotyczą jednak obiektów, których charakter nie jest z góry znany.

Teza rozprawy jest dwuczęściowa:

Teza 1: Deskryptor punktu charakterystycznego obrazu zawierający dane o położeniu punktu względem krawędzi obiektu oraz dane o głębi pozwala zwiększyć skuteczność dopasowania punktów charakterystycznych.

Teza 2: Usunięcie z procedury dopasowania punktów charakterystycznych punktów znajdujących się w obszarze krawędzi obiektu zwiększa powtarzalność detekcji punktów.

Tezy są sformułowane jasno i precyzyjnie, i odpowiadają temu co wykazano w ramach badań. Są one bardzo szczegółowe i nie pokrywają całego zakresu osiągnięć doktoranta. Rozprawa ma charakter konstrukcyjno-doświadczalny: zaproponowano i zaimplementowano nowe algorytmy związane z dopasowywaniem punktów charakterystycznych obrazów. Efektywność jednego z algorytmów porównano z efektywnością dotychczasowych rozwiązań dla starannie przygotowanego stanowiska pomiarowego.

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 przekonujący?

Bibliografia rozprawy liczy 64 tytuły, przy czym pozycje [13] i [56] można by połączyć. Prac własnych doktoranta jest 5, wszystkie są wspólne z przynajmniej jednym z promotorów. Nie jest to długa lista, niemniej wszystkie prace kluczowe są bardzo aktualne. Dobór cytowań wpisuje się w ogólną koncepcję "krótkiej" rozprawy doktorskiej, czyli zawierającej jedynie niezbędne minimum informacji o dziedzinie i badaniach autora. Wnioski doktoranta z przeglądu źródeł sprowadzają się do opisu istniejących algorytmów rozpoznawania i dopasowywania punktów charakterystycznych, metod ich badania, oraz konstrukcji stanowiska doświadczalnego. Jak więc widać, wnioski są konkretne i jasno sformułowane.

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

W rozprawie nie postawiono w sposób jawny celu pracy, co jest złamaniem ogólnie przyjętej konwencji. Rozprawa zawiera podrozdział "Motywacja", w którym przedstawiono zakres zrealizowanych badań, bardziej szczegółowo opisanych w dalszej części pracy. Ujmijmy odpowiedź na postawione pytanie następująco: rozprawa mieści się w konwencji innych prac naukowych z tej dziedziny, zawiera nowe wyniki, a sposób w jaki porównano opracowane algorytmy z istniejącymi dotychczas nie wzbudza zastrzeżeń.

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?

Oryginalny wkład doktoranta do nauki sprowadza się do opracowania dwóch nowych metod:

1. Algorytmu określanego skrótem DBFD (Depth-Based Feature Descriptor), który cechuje się zarówno zwiększoną powtarzalnością znajdowania punktów charakterystycznych przy różnych warunkach rejestracji obrazów, jak i poprawioną skutecznością dopasowania punktów charakterystycznych w obrazach 2.5D, rozdział 4, rys. 4.2 i 4.5.
2. Techniki oceny metod detekcji punktów charakterystycznych bez użycia obrazów odniesienia (ground truth), rozdział 3.3, rys. 3.4. Została ona wykorzystana do porównania istniejących metod.

Tezy rozprawy związane są z technikami użytymi w algorytmie DBDF, tak więc w rozprawie i jej streszczeniu po polsku doktorant podaje trzy główne osiągnięcia rozprawy, dwa związane z tym algorytmem. Zaproponowane rozwiązania są nowatorskie w skali światowej i mają szanse być wykorzystane w praktyce.

5. Czy Autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników /zwięzłość, jasność, poprawność redakcyjną rozprawy/?

Ogólna redakcja rozprawy jest poprawna, jest ona napisana bardzo zwięźle, czasem nawet za bardzo. Prezentacja wyników jest zrobiona jasno i przekonująco, niemniej jej tekst czyta się trudno. Autor ma niechęć do tłumaczenia, zdarzają mu się niestarannie sformułowane wypowiedzi, no i angielski pracy nie jest najlepszej próby. Układ pracy jest klasyczny: Rozdział pierwszy to motywacja badań, tezy i krótkie przedstawienie treści, rozdział drugi zawiera przegląd dotychczasowych technik detekcji punktów charakterystycznych obrazów, rozdział trzeci prezentuje metody oceny tych technik, w tym nową metodę bez obrazów odniesienia, rozdział czwarty to omówienie nowego algorytmu detekcji i dopasowania punktów charakterystycznych obrazów, oraz zastosowanej metody oceny efektywności tego

rodzaju algorytmów, w rozdziale piątym przedstawiono wyniki porównań nowego algorytmu z dwiema najważniejszymi technikami znanymi z literatury, rozdział szósty to podsumowanie rozprawy.

Redakcję tekstu i język rozprawy określiłbym jako ledwie dostateczne. Ocena wynika z dwóch istotnych uchybień: Po pierwsze, autor wielokrotnie używa ścisłych terminów bez ich objaśniania. Fatalny pod tym względem jest rozdział drugi, właściwie w opisie każdej przedstawionej w nim metody można znaleźć takie uchybienie. Chyba najgorszy pod tym względem jest opis algorytmu ORB. Pół biedy, że nie zdefiniowano co to znaczy, że współczynnik skali piramidy jest niecałkowity, oraz pojęcia centroid, gorzej, że pojawia się niewyjaśniony termin "łatka obrazu" (image patch), która ostatecznie okazuje się nieokreśloną funkcją. Funkcja ta podlega binaryzacji, po czym obliczane jest n sum ważonych z wartości zbinaryzowanej funkcji, nie wiadomo w jakim zakresie, a co najważniejsze, nie wiadomo czym te sumy się od siebie różnią. W podrozdziale opisującym detektor narożników Harrisa niezdefiniowane są pochodne cząstkowe obrazu, okno gaussowskie i nie wiadomo na jakiej zasadzie przebiegają operacje we wzorze (2.1). Komentarz do wzoru (2.7) w metodzie SURF byłby zrozumiały, gdyby wcześniej wyjaśniono jak konstruuje się piramidę Gaussowska, dalej mamy nieobjaśnione terminy "box filtering" i "sliding orientation window", wzór (2.8) też powinien być lepiej opisany. W metodzie FAST pojawia się niewyjaśnione pojęcie "Bresenham circle" itp, itd. Mam ogólną uwagę do doktoranta: zasada jest taka, że albo objaśniamy wprowadzane wzory i terminy, albo ich nie używamy i radzimy sobie w inny sposób.

W dalszej części tekstu autor utrzymuje, że w jednym z podrozdziałów przedstawił metodę SIFT. Rozpoczyna się on od opisu piramidy gaussowskiej w sposób niemożliwy do zrozumienia. Potem pojawia się bez określenia termin "hesjan". Definicję hesjana można za to znaleźć w rozdziale opisującym metodę SURF, ale sformułowaną w sposób sugerujący, że jest to szczególny hesjan ("this Hessian"). Dalej opisane są w sposób zrozumiały składowe lokalnego gradientu, ale ich użycie w celu zdefiniowania deskryptorów jest sformułowane bardzo niejasno. W żaden sposób nie można tego tekstu określić opisem metody SIFT. Podsumowując, z przeglądu algorytmów wyznaczania punktów charakterystycznych obrazów i ich deskryptorów najcenniejsze są rysunki 2.1 i 2.3, wartość informacyjna pozostałej części rozdziału drugiego jest bardzo ograniczona.

Liczne nieformalności rozdziału drugiego można jeszcze tolerować, ostatecznie praca doktorska to nie jest podręcznik. Gorzej, że podobne uchybienie pojawia się przy opisie nowej metody oceny technik detekcji punktów charakterystycznych obrazów w podrozdziale 3.3. Pod wzorami (3.6) i (3.7) stwierdza się, że obrót obrazów nimi opisany pozwala uniknąć użycia obrazu odniesienia (reference image). Niemniej dwa zdania dalej pojawia się termin "piksel odniesienia" (reference pixel), który to termin nie zostaje nigdzie dalej określony. Oczekuję, że w odpowiedzi na moją ocenę pracy doktorant przedstawi prawidłowo sformułowaną definicję tego pojęcia.

Drugie istotne uchybienie rozprawy to jej język angielski. Nawet przy mojej ograniczonej kompetencji w tym zakresie widzę, że tekst roi się od błędów językowych: nieprawidłowo użyte przyimki, źle dobrane słowa, błędy gramatyczne, jak na przykład zdania złożone zgodnie z polską składnią, mieszanie czasów w jednym zdaniu, mniej więcej dwa razy za dużo użytych rodzajników "the"... To wszystko nakłada się na skłonność do niestarannego formułowania myśli, widoczną także w polskim streszczeniu pracy. Jest zwyczaj, że recenzent wskazuje konkretne błędy językowe rozprawy, ale w tym przypadku jest to bezsensowne - samych nadmiarowych "the" jest po kilka na każdej stronie. Osobiście doktoranta nie rozumiem, doktorat jest czymś wyjątkowym, a błędy językowe są łatwo usuwalne - nie można było poprosić o pomoc anglisty?

Z pomniejszych błędów redakcyjnych: Zdjęcie sceny na rys. 1.1 zawiera puszkę farby (paint), a nie obraz (painting), rysunek 3.2 nie zgadza się z komentarzem odnośnie punktów P_3' i P_3'' , w komentarzu do wzoru (3.11) użyto nieobjaśnionego terminu Iverson bracket, na stronie 35 nie wyjaśniono o jakie interpolowane fotografie chodzi, na stronie 39 pojawia się niezdefiniowany filtr gaussowski, procedura której ilustracją jest rysunek 4.10 nie jest zbyt zrozumiała, rysunek 4.11 powtarza się dwa razy, rysunek 5.7 nie jest skomentowany w tekście pracy, są dwa różne odwołania do detektora brzegów Canny'ego, pozycja literaturowa [13] jest zbędna, co oznacza strona 260 w pozycji literaturowej [43]?

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

Główne wady rozprawy są związane z jej stroną redakcyjną i zostały opisane w poprzednim punkcie. Jeśli chodzi o stronę merytoryczną pracy to nie mam wielu zastrzeżeń i dlatego mimo obiekcji zdecydowałem się dopuścić ją do publicznej obrony. Definicja punktu charakterystycznego (keypoint) na stronie 12 jest w oczywisty sposób zbyt obszerna. Wyobraźmy sobie zdjęcie głowy konia z charakterystyczną białą "strzałką" na czole, w myśl wspomnianej definicji strzałka ta byłaby punktem charakterystycznym zdjęcia, a nie jest to obiekt punktowy. Nowa metoda oceny technik detekcji punktów charakterystycznych z rozdziału 3.3 dotyczy wyłącznie badania ich wrażliwości na obrót, tak przynajmniej wynika z treści rozdziału. Ciekawe jak zadziałałaby gdyby równania (3.6) i (3.7) zastąpiono przekształceniem afinicznym, albo bardziej ogólną formułą homograficzną. Rewelacyjny wynik algorytmu DBFD dotyczący błędnie pozytywnych rezultatów na rysunku 5.6, jest godny uwagi, niemniej czy na pewno jest to wersja optymalna algorytmu? Dla poprawy wyniku na rysunku 5.13 metodę specjalnie dostrojono. Rozważania na ten temat towarzyszące rysunkowi 4.10 sprawiają wrażenie intuicyjnych, a więc niezbyt przekonujących.

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Rozprawa zawiera opis bardzo interesującego algorytmu detekcji i dopasowania punktów charakterystycznych obrazu. Algorytm ten jest lepszy od dotychczas uważanej za wzór metody SIFT, co więcej, może być uważany za jej rozwinięcie. Te dwie cechy niemal gwarantują, że algorytm ten zostanie zauważony przez środowisko naukowe i posłuży jako baza dla dalszego rozwoju dziedziny. Nowa metoda oceny technik detekcji punktów charakterystycznych jest ciekawa przede wszystkim ze względu na unikanie definiowania obrazu odniesienia, co jest istotną nowością. Tak więc przydatność rozprawy dla nauk technicznych jest bezdyskusyjna.

8. Do której kategorii Recenzent zalicza rozprawę:

Rozprawę zaliczam do kategorii:

spełniająca wymagania stawiane rozprawom doktorskim przez obowiązujące przepisy.

W związku z powyższym wnoszę o dopuszczenie rozprawy mgr inż. Karola Matusiaka do publicznej obrony przed Radą Wydziału Elektrotechniki, Elektroniki, Informatyki i Automatyki Politechniki Łódzkiej.

R. Stank