

dr hab. inż. Leszek J. Chmielewski, prof. SGGW
Instytut Informatyki Technicznej
Szkoła Główna Gospodarstwa Wiejskiego
ul. Nowoursynowska 159, 02-776 Warszawa

Recenzja rozprawy doktorskiej

autor rozprawy:

mgr inż. Aneta Bera

Wydział Informatyki, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

tytuł rozprawy:

Szybka ekstrakcja cech obrotowo niezmienniczych za pomocą obrazów całkowych w zadaniach detekcji

Recenzja została przygotowana w odpowiedzi na powołanie na recenzenta przez Senat Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie zgodnie z uchwałą z dnia 25 stycznia 2021 r., w związku z przewodem doktorskim Pani mgr inż. Anety Bery, prowadzonym w dyscyplinie informatyka techniczna. Promotorem rozprawy jest dr hab. inż. Przemysław Klęsk, prof. ZUT.

1 Zawartość rozprawy

1.1 Omówienie ogólne

Recenzowana praca dotyczy wybranych aspektów koncepcji pomocniczych obrazów całkowych stosowanych przy obliczaniu momentów Zernike'a. Mając te pomocnicze obrazy można obliczyć momenty Zernike'a w czasie stałym, czyli niezależnym od rozmiaru okna, w jakim momenty te są obliczane (okna są kwadratowe). Zasadniczym problemem, który rozwiązano w rozprawie, jest umożliwienie obliczania pomocniczych obrazów w jednym, globalnym układzie współrzędnych kartezjańskich, podczas gdy momenty Zernike'a są zdefiniowane w lokalnych układach współrzędnych biegunowych. Momenty te umożliwiają obliczenie cech obrazu mających właściwość niezmienniczości względem obrotu i skali. Pomocnicze dane przyspieszają obliczenia, jeśli liczy się momenty dla wielu okien w obrazie, co ma miejsce w praktycznych zastosowaniach. Koncepcja ta została opisana w pracy cytowanej w recenzowanej rozprawie jako (Bera, Klęsk i Sychel, 2019), opublikowanej w *IEEE Trans. PAMI*. Ta część rozprawy doktorskiej ma charakter teoretyczny, zaś pozostałe części – praktyczny.

W pracy zaproponowano trzy dalsze metody udoskonalające w praktyce zastosowanie metody momentów Zernike'a. Pierwsza z nich polega na obliczaniu obrazów pomocniczych z pierścieniowych fragmentów obrazów źródłowych, co daje możliwość zwiększenia liczby uzyskiwanych cech. Metoda ta okazała się skuteczna jedynie dla obiektów w obrazie mających proste kształty. Została ona opublikowana również w pracy (Bera, Klęsk i Sychel, 2019) w *IEEE Trans. PAMI*.

Druga metoda dotyczy problemu błędów numerycznych wynikających tego, że różnice rzędów wielkości danych w obrazach pomocniczych w ich kolejnych fragmentach są bardzo wielkie, z powodu występowania wysokich potęg współrzędnych obrazu we wzorach. Dlatego ich reprezentacja w językach programowania jest niewystarczająca dla objęcia zakresu ich wartości, co

powoduje utratę części informacji. Silnie ujawnia się to w różnicach, wykorzystywanych właśnie w obliczeniach cech obrazów. Sposobem na to jest podział danych pomocniczych na obszary, w których dane nie są zniekształcane w używanej reprezentacji. Metodę tę zaprezentowano w publikacji współautorskiej z udziałem Doktorantki cytowanej w pracy jako (Kłęsk, Bera i Sychel, 2020), opublikowanej w *Computational Science, Proc. ICCS*.

Trzecie udoskonalenie polega na zastosowaniu i oprogramowaniu rozszerzenia zbioru momentów Zernike'a, a co za tym idzie, liczby cech które można policzyć dla obrazu. Rozszerzenie takie było już sugerowane w literaturze, natomiast w pracy zostało zrealizowane.

Jako mechanizm klasyfikacyjny zastosowano drzewa decyzyjne połączone z kosztami z odpowiedzią rzeczywistoliczbową. Utworzone w ten sposób słabe klasyfikatory zintegrowano za pomocą algorytmu RealBoost. Klasyfikatory stosowano właściwie jako detektory obiektów, klasyfikując podobszary obrazu jako należące do obiektu lub do tła.

Metody przebadano przeprowadzając dwa eksperymenty, jeden na obrazach zawierających prostsze obiekty, a drugi bardziej skomplikowane. Z wyników można wywnioskować, że efekty detekcji są korzystne.

Praca składa się z wprowadzenia, ośmiu rozdziałów, spisów rysunków, tablic i algorytmów oraz bibliografii.

We wprowadzeniu zawarto między innymi opis celu i tezy pracy. Rozdział pierwszy zawiera teorię dotyczącą pomocniczych obrazów całkowych oraz opis procedury detekcyjnej.

W drugim rozdziale opisano teorię momentów Zernike'a, kwestię ich niezmienniczości ze względu na skalowanie oraz niezmienniczość ich modułów ze względu na obrót. Pokazano to wyprowadzając odpowiednie wzory. Dokonano obszernego porównania momentów Zernike'a (ZM) z ortogonalnymi momentami Fouriera-Mellina (OFMM), które z kilku powodów mogłyby być uważane za lepsze deskryptory obiektów, lecz nie można ich szybko obliczać za pomocą pomocniczych obrazów całkowych. Z analizy literatury wynika, że inne momenty używane w podobnych zastosowaniach mają mniej korzystne właściwości.

W rozdziale trzecim opisano kwestie stałoczasowego obliczania momentów Zernike'a. Dokładnie przeanalizowano efektywność algorytmów obliczania całkowych obrazów pomocniczych. Zaprezentowano też koncepcję liczenia cech z pierścieni, zamiast z kwadratowych okien w obrazie, dla zwiększenia liczby uzyskiwanych cech.

Koncepcje te zweryfikowano w rozdziale czwartym na dwóch zbiorach danych obrazowych. Pierwszy zawierał tylko literę „A” pisaną różnymi czcionkami na różnych tłach i uzyskano dla niego dobre wyniki (na danych testowych czułość 0.98, specyficzność $1 - 2 \cdot 10^{-6}$). Drugi zawierał samoloty na zdjęciach rzeczywistych i wyniki uzyskane dla niego zostały przez Autorkę uznane za negatywne (choć na danych walidacyjnych uzyskano czułość około 0.9 przy specyficzności 0.99), zatem w tej koncepcji wyników dla danych testowych nie podano.

W rozdziale piątym zaprezentowano koncepcję podziału obrazów pomocniczych na *kawalki*, dla zaradzenia wspomnianego przekroczenia zakresu wartości możliwego w reprezentacji zmiennych w języku programowania (stosowano typ zmiennoprzecinkowy podwójnej precyzji). Rozdział zamyka analiza nakładu obliczeniowego na obsługę podzielonych obrazów, który nie przekracza 50% wielkości tego nakładu bez podziału.

Rozdział szósty jest poświęcony eksperymentom z opisaną metodą podziału obrazów pomocniczych, z tymi samymi zbiorami obrazów, co poprzednio. Pokazano, że uzyskiwane wyniki są wyraźnie lepsze, niż w metodzie bez podziału. Na zbiorze „A”, podzbiór testowy, uzyskano czułość 0.992 przy specyficzności wynoszącej 1.00 (brak fałszywie pozytywnych rozpoznań). Na zbiorze samolotów uzyskano czułość 0.706 przy specyficzności $1 - 2.8 \cdot 10^{-6}$ (dla porównania

z wynikami bez podziału na kawałki, na danych walidacyjnych z podziałem czułość 0.9 uzyskano dla specyficzności około $1 - 2 \cdot 10^{-5}$).

W podsumowaniu Autorka uczciwie przyznaje, że miary jakości detekcji obiektów nie są lepsze, niż przy zastosowaniu cech Haara lub deskryptora HOG (*Histogram of Oriented Gradients*), co wynika z tego, że informacja zawarta w używanych do detekcji momentach Zernike'a niezbyt wysokich rzędów nie zawiera dostatecznie dokładnej reprezentacji kształtu obiektów. Remedium byłoby zastosowanie momentów znacznie wyższych rzędów, co skomplikowałoby kwestię dokładności numerycznej. Gdyby zastosować zmienne poczwórnej precyzji, te komplikacje znikłyby. Dlatego zdaniem Doktorantki warto pracować nad cechami opartymi na momentach Zernike'a, które ze względu na swą niezmienniczość względem obrotu i szybkość obliczania są atrakcyjną alternatywą dla innych cech obiektów w obrazie.

Bibliografia kończąca rozprawę zawiera około 80 pozycji, w tym jedna z roku 2021, pięć z 2020, dwie z 2019. Doktorantka jest współautorem dwóch z cytowanych publikacji, za 200 i 140 punktów.

1.2 Cel i tezy pracy

Cel pracy został sformułowany we wprowadzeniu:

- **Celem pracy jest opracowanie algorytmu umożliwiającego szybkie (stałoczasowe) obliczanie momentów Zernike'a w ramach procedury detekcyjnej, niezależniając się od pozycji okna i liczby pikseli w oknie.**

Postawiono następującą tezę:

- **Po przygotowaniu odpowiedniego zestawu zespolonych obrazów całkowych przed procedurą detekcyjną, momenty Zernike'a dla każdego kwadratowego okna obrazu mogą zostać wyznaczone w czasie stałym – $O(1)$ – niezależnie od liczby pikseli w oknie.**

2 Omówienie treści i wyników rozprawy

2.1 Uwagi pozytywne

Wartość tezy pracy Reprezentacja momentów Zernike'a, zdefiniowanych pierwotnie w układzie biegunowym, w postaci umożliwiającej skorzystanie z danych pośrednich, zdefiniowanych w układzie kartezjańskim, które są jednolite dla całego obrazu, jest oryginalnym i wartościowym osiągnięciem. Jego oryginalność została potwierdzona publikacją w czasopiśmie *IEEE Trans. PAMI*, o wysokim prestiżu (200 punktów MNiSzW). Praca ta została również wyróżniona w Konkursie Oddziału PAN w Gdańsku dla młodych naukowców. Zwraca uwagę łatwość, z jaką Doktorantka przekształca skomplikowane wzory matematyczne. Jest to pozytywna cecha w czasach, gdy dominują podejścia eksperymentalne bazujące na symulacjach komputerowych.

Wartość pozostałych koncepcji udoskonalających metodę z tezy Wprowadzenie obliczania obrazów pomocniczych wymagała solidnej analizy wzorów, dość skomplikowanych. Wprawdzie zaleca się, aby badane okno obrazu mieściło się całkowicie w jednym kawałku, to

jednak oprogramowano również wszystkie inne, mniej korzystne przypadki. Podano oszacowanie dla liczby dodatkowych operacji oparte na solidnie wyprowadzonych wzorach, nie pomijając także podejścia empirycznego (otrzymano zgodne wartości).

Koncepcja liczenia cech tylko z pierścieni jest ciekawa, i choć nie okazała się skuteczna gdy obiekty są skomplikowane, to można ją stosować w innych przypadkach.

Zwiększanie liczby cech opartych na momentach Zernike'a okazało się skuteczne. Również wymagało biegłości matematycznej i rozwiniętych umiejętności porządkowania wzorów.

Koncepcje te zostały również potwierdzone publikacjami, częściowo w czasopiśmie *IEEE Trans. PAMI*, i częściowo w *Computational Science, Proc. ICCS* (140 punktów MNiSzW).

Dobry przegląd literatury Praca jest solidnie osadzona w kontekście literatury dotyczącej podobnych zagadnień. Literatura jest nie tylko zebrana i wzmiankowana, ale także wystarczająco jasno napisano, co istotnego dla problemów opisywanych w rozprawie zawierają kolejne pozycje.

Wysokie wymagania dla miar jakości wyników Uzyskiwane wyniki detekcji mają dokładność przekraczającą 0.99 (dane testowe: Tabl. 4.3, 7.2, zbiór „A”, Tabl. 7.4, zbiór „Samoloty”), zaś wyniki dla danych walidacyjnych, zbiór „Samoloty”, gdzie czułość przekracza 0.9 przy specyficzności 0.99 są potraktowane jako *bardzo słabe* (prawdopodobnie chodzi raczej o to, że jakość na danych testowych była znacznie gorsza, i dlatego jej nie ujawniono). W wielu badaniach takie wyniki byłyby określone jako całkiem dobre, co zamknęłoby drogę do poszukiwania lepszych rozwiązań.

Ilustracja tłumienia informacji blisko środka obiektu W ciekawy i poglądowy sposób przedstawiono to zjawisko dla momentów Zernike'a, ortogonalnych momentów Fouriera–Mellina i momentów zespolonych (Rozdz. 2.5.2, Rys. 2.5). Skorzystano z koncepcji przedstawionej w literaturze (Abu-Mostafa i Psaltis, 1984), ale realizacja tego przedstawienia jest przekonująca.

Bardzo wysoki poziom edycyjny rozprawy Poziom ten ujawnia się szczególnie w staranności i wiedzy o zasadach składania wzorów matematycznych. Jest to bardzo cenne, gdyż ujawnia szacunek dla treści. Zastosowanie tak profesjonalnego narzędzia do składu publikacji, jakim jest \LaTeX , nie ułatwia jeszcze wszystkiego; Doktorantka wykazała się także wiedzą.

2.2 Uwagi dyskusyjne i krytyczne

2.2.1 Uwagi merytoryczne

Moje uwagi merytoryczne są nieliczne i należy je traktować raczej jako okazję do dyskusji niż krytykę.

Brak bezpośredniego odniesienia do niezmienniczości cech Ta uwaga odnosi się przede wszystkim do niezmienniczości względem skali, ale w pewnym zakresie także do niezmienniczości względem obrotu. Co do skali, to stosuje się analizę wieloskalową (okna o stopniowo zmieniającej się wielkości), podczas gdy cechy zapewniają w tym zakresie niezmienniczość. Co do obrotu, to odnalazłem jedyne miejsce, gdzie niemal jawnie korzysta się z niezmienniczości, tam gdzie wskazano, że obiekty uczące w zbiorze „A” były obracane w zakresie jedynie od -45° do

45°, a nie w całym zakresie pełnego obrotu. Mimo to, uzyskano dobre wyniki detekcji liter obróconych o dowolny kąt.

Przypuszczam, że sprawę tych dwóch niezmienniczości potraktowano w powyższy sposób świadomie, ale chciałbym usłyszeć komentarz Doktorantki na ten temat.

Brak odniesienia do skutków tłumienia, szumów i drgań momentów Zernike'a w eksperymentalnej części pracy Wspomniany już ciekawy opis zjawiska tłumienia informacji w centralnej części obrazu (Rozdz. 2.5.2, Rys. 2.5) i ilustracja niedokładnej reprezentacji obiektów za pomocą nawet dużej liczby momentów (Rozdz. 2.3, Rys. 2.1) wzbudza u czytelnika zaniepokojenie co do tego, jak w pracy poradzono sobie z tymi problemami. Jednak w dalszej części pracy brak jawnego odniesienia do tych problemów.

2.2.2 Drobne uwagi redakcyjne i techniczne

Uwagi te dotyczą spraw redakcyjnych, które są merytorycznie nieistotne i nie wpływają na ocenę pracy. Zamieszczam je do wiadomości Autorki, która w przeważającej części tekstu złożyła pracę z ogromną dbałością. Omawiam je w kolejności stron.

- S. 9¹⁰: HOG – wszyscy znają ten skrótowiec, ale wyjaśnić trzeba przy pierwszym użyciu (*Histogram of Oriented Gradients*).
- S. 9₄: *W ostatnich czasach (...) 1984 (...) 1995...* : te lata, to być może są *ostatnie czasy* dla mnie, ale chyba nie dla Doktorantki (w sensie *niedawne*).
- S. 13₁₄: *Celem pracy jest opracowanie (...) uniezależniając się...* : wszechobecny błąd, *anakolut*. Co jest celem, a kto się uniezależnia? Anakolut znaleziono również na s. 32¹³⁻¹⁴: *zmieniając kolejność całkowania (...), momenty mogą zostać wyrażone...* . To nie momenty zmieniają kolejność całkowania. Kolejny jest na s. 16₆₋₅: *Obraz całkowity może zostać wyznaczony (...), wykorzystując wzory...* , dalsze na ss. 20¹⁴⁻¹⁵, 20¹⁸⁻¹⁹, 41₁₂₋₁₀ itd.
- S. 16¹: Wiersz zawieszony (bękart), ostatni wiersz akapitu samotnie otwierający nową stronę.
- S. 17₂: Należałoby wyjaśnić oznaczenia w zapisie $(x_1, y_1), \dots, (x_m, y_m)$. Domyślamy się, że x_i to wektor cech, a y_i to indeks klasy obiektu uczącego i , $i = 1, \dots, m$. Jednak w opisie Algorytmu 1 (s. 15) zmienna x oznaczała współrzędną poziomą obrazu, zaś z kolei w Algorytmie 2 (s. 18) okazuje się, że zmienna y_i nie jest zwykłym indeksem klasy, bo należy do zbioru $\{-1, +1\}$.
- S. 18 i dalsze: Algorytm RealBoost bywa nazywany Real AdaBoost – niejednolitość nomenklatury.
- S. 20₁₁: *(...) gdzie m_1 i m_2 to liczba przykładów w danym podzbiornie \rightarrow (...) gdzie m_1 i m_2 to liczby obiektów należących do podzbiornów D_1, D_2 , odpowiednio.*
- S. 20₆₋₅: *Proces powtarza się tak długo, aż osiągnięty węzeł nie jest liściem.* To znaczy, kończymy gdy jest liściem, czy gdy nim nie jest?
- S. 21¹⁰: Słowa *decision stumps* jako obce należałoby pisać itałikiem.

- S. 21, tekst do wzorów (1.12-1.14): Nie wyjaśniono pojęć *wag* (utożsamianych z *prawdopodobieństwami*) ani *rund*.
- S. 23₅: Nie *rozmiar* zbioru, lecz jego *moc*.
- S. 30, Rys. 2.2, 2.3: Należałoby pokazać używaną skalę barwną.
- S. 31, Rys. 2.4: Trudno dobrać barwy tak, aby się wystarczająco różniły. Bardzo polecam <http://www.ColorBrewer.org>.
- 51₆: Jest trochę nieściśle, bo pokazano nie same indeksy, ale wartości tablicy.
- S. 58, Rys. 4.2 i Tab. 4.2: Oznaczenia [8, 8, 8] użyte na rysunku są wyjaśnione dopiero dalej, w tabeli. Również, nie wiem, jak się mają wartości czułości z tabeli, wynoszące .964 i .951, skoro na odpowiednich wykresach na rysunku w ogóle nie ma danych dla tak małych wartości?
- S. 59, Rys. 4.3: W trzecim rzędzie na żadnym obrazie nie jest wykryta litera P.
- S. 61, Rys. 4.4: Rysunek (c) zawiera przecież fragmenty samolotów, a nawet jeden prawie cały samolot (nieostry), więc są to raczej trudne pozytywy.
- S. 63¹⁴: Po co pisać, że najmniej znaczące cyfry to skrajnie prawe cyfry?
- S. 65⁶: Zamiast pisać w przenośni „*prywatny*” układ współrzędnych, lepiej napisać ściśle *lokalny*.
- S. 73₉: Nie Znak stopnia kąтового wygodnie jest pisać tak: 45° .
- S. 96, Rozdz. 8.1: Rozdział 8 ma wydzielony jeden podrozdział. Jest to błąd edycyjny; problematyczny jest podział na jedną część. Należałoby zastosować konstrukcję jak we Wprowadzeniu, w którym wydzielono nienumerowaną część zawierającą cel i tezę pracy.
- Bibliografia:** Zamieszczenie identyfikatorów DOI pomogłoby wyszukiwać publikacje w sieci.
- Całość pracy:** Byłoby korzystne użycie w składzie tekstu w MiKTeXu pakietu *hyperref*. Bez stosowania dodatkowych zabiegów w wersji elektronicznej rozprawy byłyby wtedy obecne linki do wzorów, tabel, rysunków i przede wszystkim pozycji literatury.

3 Podsumowanie

W rozprawie można wyróżnić istotne elementy oryginalne oraz inne pozytywne aspekty omówione w rozdziale 2.1, zaś uwagi dyskusyjne i krytyczne omówione w rozdziale 2.2 nie umniejszają w istotny sposób wartości pracy.

Cel pracy został osiągnięty, a teza została wykazana.

W rozprawie zaprezentowano oryginalną, wartościową i ciekawą metodę przyspieszającą obliczenia cech w których wykorzystano momenty Zernike’a, przy czym osiągnięto czas stały, niezależny od wielkości okna. Metodę tę głęboko rozpracowano, opierając się w znacznym zakresie na uzasadnieniach matematycznych, oraz uzupełniono ją dodatkowymi usprawnieniami. Całość zweryfikowano na przykładach praktycznych.

Wnioski Powyższy opis, uwzględniający uwagi pozytywne jak również uwagi dyskusyjne i krytyczne, uzasadnia moje ostateczne wnioski o następującej treści.

Recenzowana rozprawa w postaci opracowania pisemnego **stanowi oryginalne rozwiązanie problemu naukowego** w dyscyplinie informatyka techniczna i telekomunikacja. Rozprawa **prezentuje ogólną wiedzę teoretyczną kandydatki w tej dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej.** Zagadnienie badawcze zostało prawidłowo postawione i skutecznie rozwiązane, a rozwiązanie zostało rzetelnie zweryfikowane. Tym samym rozprawa spełnia wymagania obowiązującego prawa w zakresie rozpraw doktorskich. Rozprawę oceniam pozytywnie i stawiam wniosek o skierowanie jej do dalszych etapów przewodu doktorskiego.



