

**RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ**
zaprasza na

**PUBLICZNĄ OBRONĘ ROZPRAWY DOKTORSKIEJ
mgr. inż. Kacpra Sarnackiego**

która odbędzie się w dniu **7 kwietnia 2022 roku o godzinie 12:00** w trybie hybrydowym

Temat rozprawy doktorskiej:

„Poprawa jakości zdjęć pisma ręcznego z wykorzystaniem morfologii adaptacyjnej”

Promotor: prof. dr hab. inż. Khalid Saeed - Politechnika Białostocka

Recenzenci: prof. dr hab. inż. Krzysztof Cpałka – Politechnika Częstochowska

prof. dr hab. inż. Sławomir Wierzchoń - Instytut Podstaw Informatyki PAN w Warszawie

prof. dr hab. inż. Michał Woźniak – Politechnika Wrocławska

Obrona odbędzie się w budynku Wydziału Matematyki i Nauk Informatycznych Politechniki Warszawskiej (Koszykowa 75, 00-662 Warszawa) w sali nr 40 oraz z wykorzystaniem platformy MS Teams. Osoby zainteresowane uczestnictwem w obronie proszone są o zgłoszenie chęci uczestnictwa w formie elektronicznej na adres sekretarza komisji: dr hab. inż. Maciej Grzenda, prof. uczelni – email : m.grzenda@mini.pw.edu.pl do dnia 4 kwietnia 2022 do godz. 13:00., zaznaczając w zgłoszeniu preferowaną formę uczestnictwa (stacjonarną lub zdalną). W przypadku wyczerpania limitu miejsc dozwolonych przez regulacje Politechniki Warszawskiej do wykorzystania w trybie stacjonarnym, rezerwujemy sobie prawo poinformowania o konieczności skorzystania z udziału zdalnego.

Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Warszawa, Plac Politechniki 1.

Streszczenie rozprawy doktorskiej i recenzje są zamieszczone na stronie internetowej: <https://bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego/Doktoraty/Wszczete-po-30-kwietnia-2019-r/Dyscyplina-informatyka-techniczna-i-telekomunikacja-dziedzina-nauk-inzynierjno-technicznych/mgr-inz.-Kacper-Sarnacki>.

Przewodniczący Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej

dr hab. inż. Jarosław Arabas

Streszczenie

Niniejsza praca przedstawia algorytm usprawniający jakość zdjęć pisma ręcznego z wykorzystaniem morfologii adaptacyjnej. Celem jej powstania jest udowodnienie tezy, że istnieje uniwersalne podejście do poprawiania jakości binaryzacji obrazów zawierających obiekty o liniowej strukturze, które da lepsze wyniki w porównaniu do binaryzacji. Niewątpliwą zaletą prezentowanego podejścia jest modularność. Algorytm współpracuje z dowolną metodą binaryzacji. Sposób wyznaczania pola kierunkowego może być także wybrany przez użytkownika. Ponadto prezentowany algorytm proponuje stosowanie komponentów opcjonalnych. Pozwalają one uzyskać jeszcze lepszy rezultat, jednak kosztem złożoności obliczeniowej. Przykładem takich rozszerzeń jest zwiększanie próbkowania czy test 4-sąsiedztwa. Elastyczność prezentowanej metody pozwala znaleźć balans pomiędzy jakością zwracanego rozwiązania a wydajnością działania. Wysoką skuteczność algorytm zawdzięcza identyfikacji pola kierunkowego, dzięki czemu możliwe jest przewidzenie kształtu linii czy ewentualnych przecięć. Na tej podstawie można dostosować sposób naprawiania ubytków wewnątrz linii poprzez zastosowanie różnych typów morfologii.

Szczególnie istotną częścią niniejszej pracy jest weryfikacja jakości stworzonego algorytmu. Z racji jego elastyczności, testom zostały poddane różne wersje, składające się z wymiennych komponentów. W celu obiektywnej oceny rozwiązania, zastosowano trzy rodzaje weryfikacji. Użyto testów wizualnych, analitycznych i statystycznych. Operowanie na obrazach przedstawiających pismo pozwala w łatwy sposób weryfikować uzyskane usprawnienia, dzięki powszechnie znanej wiedzy na temat kształtu liter. Test wizualny jest użyteczny także przy uwzględnieniu faktu, że w niektórych sytuacjach poprawianie rezultatów binaryzacji jest ostatnim etapem automatycznego przetwarzania obrazów, zanim tak przetworzone zdjęcie trafi do specjalisty. Kolejnym rodzajem testów jest test analityczny, który opisuje liczbowo jakość algorytmu. Dzięki niemu można porównać dokładność różnych wersji algorytmów. Ta metoda porównuje wyniki z rezultatem idealnym. Dzięki temu można ocenić, czy algorytm poprawiający jakość binaryzacji rzeczywiście polepsza ostateczny rezultat oraz w jakim stopniu różni się od wyniku idealnego. Testy statystyczne udowadniają, że poprawa jakości obrazu poprzez stosowanie prezentowanego algorytmu nie jest przypadkowa i że jest to zmiana istotna statystycznie.



INSTYTUT PODSTAW INFORMATYKI
POLSKIEJ AKADEMII NAUK
ul. Jana Kazimierza 5, 01-248 Warszawa

tel.: ++(48) 22 38-00-500
fax.: ++(48) 22 38-00-510

e-mail: ipl@ipipan.waw.pl
www.ipipan.waw.pl

Prof. dr hab. inż. Sławomir Wierzchoń
stw@ipipan.waw.pl
Profesor zwyczajny
<http://www.ipipan.waw.pl/~stw>

Recenzja rozprawy doktorskiej
mgr inż. Kacpra Sarnackiego

p.t.

*Poprawa jakości zdjęć pisma ręcznego
z wykorzystaniem morfologii adaptacyjnej*

Przedmiotem recenzji jest rozprawa doktorska Pana mgr inż. Kacpra Sarnackiego o podanym wyżej tytule, która powstała w ramach Studium Doktoranckiego na Wydziale Matematyki i Nauk Informacyjnych Politechniki Warszawskiej pod kierunkiem naukowym Pana prof. dr hab. inż. Khalida Saceda. Recenzję przygotowano na zlecenie Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja, przesłane niżej podpisanemu listem datowanym 16 listopada 2021r.

1. Przedmiot pracy

Rozprawa poświęcona jest problemowi rozpoznawania pisma ręcznego dostępnego na statycznych nośnikach, głównie na zdjęciach. Jak pisze Autor: „celem niniejszej pracy jest udowodnienie tezy, że istnieje uniwersalne podejście do preprocessingu obrazów zawierających obiekty o liniowej strukturze, które poprawiałyby wyniki zwracane przez binaryzację” (s. 12 rozprawy). Wymieniona tu *binaryzacja* to istotna operacja wstępnego przetwarzania obrazu polegająca na podziale wszystkich pikseli na dwie klasy: tę która reprezentuje tło i tę reprezentującą pismo. Autor proponuje dość ogólną (nazywa ją „uniwersalną”) procedurę realizującą postawiony wyżej cel.

2. Układ i zawartość pracy

Recenzowana praca liczy 92 strony i składa się z sześciu rozdziałów oraz bibliografii zawierającej 109 pozycji. Na wstępie (nienumerowany dwustronicowy rozdział) Autor charakteryzuje istotę swojej idei, która wykorzystuje morfologię sterowaną polem kierunkowym do doboru elementów strukturalnych dla konkretnych fragmentów analizowanego obrazu.

Rozdział drugi to przegląd literatury z zakresu rozpoznawania pisma i obiektów o strukturze liniowej.

Rozdział trzeci to opis autorskiego podejścia do *preprocessingu*. Składa się on z 6 podrozdziałów, które ze względu na ich długość będą nazywał punktami. W punkcie 3.1 przedstawiono ogólny opis proponowanej metody (sposobu postępowania). W istocie jest to 3-punktowa recepta, zgodnie z którą: (a) wyznacza się pole kierunkowe dla zbinaryzowanego obrazu, (b) stosuje się operacje morfologiczne sterowane wyznaczonym polem kierunkowym, a wreszcie (c) usuwa niewielkie artefakty wykorzystując test 4-sąsiedztwa. W punkcie 3.2 przedstawiono trzy zasadnicze metody binaryzacji (metodę Osu, Sauvola i metodę Niblacka). Omówiono ich wady i zalety oraz podano wskazówki literaturowe. W punkcie 3.3 opisano podstawowe operacje morfologiczne, tzn. dylatację i erozję, a następnie wspomina się o ich złożeniach, czyli otwarciu morfologicznym i domknięciu morfologicznym. Punkt ten konkluduje Autor stwierdzeniem „ (...) żadna z przedstawionych metod nie jest odporna zarówno na problem przerwania linii pierwotnie połączonej jak i problem zakrywania luk, które są ważną częścią obiektu” (s. 48 rozprawy). Dlatego w kolejnym, czwartym, punkcie wprowadza się – grające istotną rolę w koncepcji Autora – pojęcie pola kierunkowego. Jak motywuje Autor „Autorskie rozwiązanie nie narzuca konkretnej metody” (s. 13), dlatego omówiono tu trzy metody wyznaczania pola kierunkowego. Opcjonalnym komponentem algorytmu jest Test 4-sąsiedztwa, który opisano w punkcie 3.5. W ostatnim punkcie tego rozdziału przedstawiono dodatkowe rozwiązanie (moduł) umożliwiające zwiększenie liczby próbek.

W czwartym rozdziale opisano miary stosowane do oceny analizowanych technik.

Rozdział 5 poświęcono wynikom przeprowadzonych eksperymentów, natomiast rozdział 6 to podsumowanie i wnioski.

3. Uwagi

Recenzowana rozprawa ma w zasadzie charakter empiryczny. Należy jednak zaznaczyć, że opisane w rozdziale 5 eksperymenty wykonano z użyciem narzędzi i algorytmów, które Autor współtworzył. Mgr Kacper Sarnacki jest współautorem czterech artykułów. Pierwszy z nich, pozycja [3] w rozprawie, ukazał się w 2021 r. w prestiżowym czasopiśmie *Information Science* (200 pktów w Wykazie Czasopism Naukowych i Recenzowanych Materiałów z Konferencji Międzynarodowych). Przedstawiono tu koncepcję procedury będącej sercem recenzowanej rozprawy (por. rozdział 3). Trzy pozostałe artykuły (pozycje [82], [83] i [105] rozprawy), publikowane w latach 2017-2019, ukazały się w materiałach dziedzicznych konferencji międzynarodowych. Jak czytamy na s. 8 rozprawy „Autor rozprawy zaimplementował algorytm i napisał cały kod, a także wykonał mnóstwo eksperymentów przeprowadzonych na różnych bazach obrazów, aby (...) udowodnić skuteczność algorytmu w praktycznych zastosowaniach”. Owa skuteczność algorytmu jest badana i potwierdzana eksperymentalnie w w/w pracach [83] i [105].

Nie mogę jednak pominąć uwag krytycznych, które nasunęły mi się w trakcie czytania rozprawy:

1. Rozprawa jest, mówiąc najogólniej, rozczarowująca jeśli chodzi o jej zawartość i sposób prezentacji materiału. Autor jest typowym reprezentantem tej młodej grupy adeptów nauki, której członkowie uważają, że skoro wszystko napisano w książkach i jest dostępne w Internecie, to po zawracać sobie głowę „przepisywaniem”. Koncentrują się oni wyłącznie na tym, co jest dla nich istotne.
2. Dokonując przeglądu piśmiennictwa Autor całkowitym milczeniem pomija istnienie ogólnodostępnych systemów konwertujących pismo ręczne na tekst. Przykładowo, Google Lens oferuje opcję łatwego kopiowania tekstów ze sfotografowanych

odręcznych notatek. Microsoft oferuje program OneNote o podobnych możliwościach. Takich programów istnieje bardzo dużo. Moim zdaniem warto było poświęcić temu tematowi chociaż skromny akapit, wskazując przy okazji specyfikę rozwiązywanego problemu.

3. Rozczarowujący jest rozdział 5, który powinien stanowić (oprócz rozdziału 3) główną część pracy. Autor kilkakrotnie podkreśla, że jego celem jest „udowodnienie tezy, że istnieje uniwersalne podejście do preprocessingu (...), które poprawiałyby wyniki zwracane przez binaryzację”. Jak rozumiem Autor chciałby „udowodnić” ową tezę drogą dużej liczby eksperymentów. Na s. 8 pisze On, że: „wykonał mnóstwo eksperymentów przeprowadzonych na różnych bazach obrazów”. Tymczasem w punkcie 5.2 mamy przedstawiony pojedynczy obraz z bazy ICFHR (rys. 34), na którym zilustrowano problemy ciągłości linii i niewielkich pętli występujące po zastosowaniu standardowych operacji morfologicznych. Niejako konkluzją tego punktu jest rysunek 45 prezentujący „najlepszy rezultat” uzyskany drogą zastosowania Autorskiej recepty. Ja osobiście nie widzę różnicy między tym rysunkiem a rysunkiem 44, na którym przedstawiono rezultaty zastosowania domknięcia morfologicznego. Może zależy to od wzroku? W punkcie 5.3 dokonano porównania miar dokładności (Acc i $Acc2$) dla wybranych metod binaryzacji i preprocessingu na obrazach z dwóch baz danych. Nie jest jednak jasne, co oznaczają wyniki przedstawione w Tabelicy 2. Czy są to wyniki uśrednione po wszystkich obrazach, czy może wyniki najlepsze/najgorsze? Jest to zdecydowanie zbyt lakoniczny opis.
4. Wprawdzie Autor pisze, że „zaimplementował algorytm i napisał cały kod”, ale nie precyzuje języka w jakim wykonano tę implementację. W takich językach jak na przykład MATLAB/Octave czy R dostępne są gotowe funkcje pozwalające realizować większość operacji, o których wspomina się w pracy.

4. Konkluzja

Biorąc powyższe pod uwagę, a w szczególności uzyskane wyniki, jak również dobór tematu, cel i zakres pracy stwierdzam, że:

- Temat rozprawy, zastosowane metody oraz uzyskane wyniki odpowiadają randze rozpraw doktorskich.
- *Rozprawa doktorska prezentuje ogólną wiedzę teoretyczną kandydata w dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej*, co potwierdzają cztery artykuły wspomniane w punkcie 3 recenzji.
- *Przedmiotem recenzowanej rozprawy jest oryginalna propozycja uniwersalnej procedury* takiej realizacji wstępnego przetwarzania obrazu, która poprawia wyniki zwracane przez binaryzację.
- Przedstawione rozwiązanie ma implikacje teoretyczne jak i praktyczne. Potencjalne kierunki eksploracji i zastosowań omawia Autor w rozdziale 6.

Stwierdzam zatem, że opiniowana rozprawa, mimo pewnych niedociągnięć, spełnia wymagania sprecyzowane w art 187 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668 z późniejszymi zmianami). Stawiam wniosek o dopuszczenie mgr inż. Kacpra Sarnackiego do publicznej obrony recenzowanej rozprawy.

Warszawa, 18 stycznia 2022 r.



Wrocław, 7 stycznia 2022r.

Prof. dr hab. inż. Michał WOŹNIAK
Wydział Informatyki i Telekomunikacji
Politechnika Wroclawska

Recenzja rozprawy doktorskiej

autor: mgr inż. Kacper SARNACKI

tytuł „Poprawa jakości zdjęć pisma ręcznego z wykorzystaniem morfologii adaptacyjnej”

Promotor: prof. dr hab. inż. Khalid Saeed

Obszar problemowy rozprawy

Recenzowana rozprawa związana jest z algorytmami poprawy jakości binaryzacji obiektów o liniowej strukturze. Otrzymane rezultaty mogą zostać wykorzystane do budowy nowych systemów analizy zdjęć pisma odręcznego z wykorzystaniem zaproponowanych metod bazujących na morfologii adaptacyjnej. Liczba obecnie realizowanych projektów oraz pozycji literaturowych związanych z algorytmami analizy obrazów dowodzi, że proponowana tematyka rozprawy jest jak najbardziej zgodna z aktualnym nurtem badań nad wizją komputerową. Cel pracy, który dotyczy opracowanie algorytmów poprawiających jakości algorytmów binaryzacji z wykorzystaniem apriorycznej wiedzy o liniowości struktur obiektów na obrazie, należy uznać za ambitny i mieszczący się w obszarze wizji komputerowej, dyscyplinie informatyka techniczna i telekomunikacja. Zaproponowana w rozprawie teza, mówiąca że stosując odpowiednie metody preprocessingu do obiektów o liniowej strukturze można polepszyć jakość ich binaryzacji, jest prawidłowa i ambitna.

Kompozycja i redakcja rozprawy

Recenzowana rozprawa składa się ze wstępu, streszczenia w języku polskim i angielskim, sześciu rozdziałów oraz spisu literatury obejmującego 109 pozycji. W rozprawie brakuje spisu oznaczeń i stosowanych skrótów oraz indeksu rzeczowego, choć umieszczanie go nie jest powszechną praktyką w przypadku rozpraw doktorskich, ale na pewno wpłynęłoby na czytelność rozprawy. Struktura pracy jest następująca:

- We wstępie przedstawiono motywację autora, cele i tezę pracy oraz syntetyczny opis zawartości rozprawy.
- Rozdział drugi zawiera krytyczny przegląd literaturowy metod preprocessingu.
- W kolejnej części przedstawiono autorski algorytm poprawy jakości binaryzacji obrazów o liniowej strukturze, który jest silnie sparametryzowany w zakresie doboru metod w kolejnych krokach przetwarzania i cechuje się wysoką modularnością.
- Rozdział czwarty opisuje metody weryfikacji zaproponowanego rozwiązania, w tym definicję oryginalnej miary jakości.
- Następny rozdział zawiera wyniki badań eksperymentalnych oraz ich analizę.
- W ostatniej części autor podsumowuje krótko najważniejsze osiągnięcia rozprawy oraz wskazuje potencjalne obszary zastosowania zaproponowanych algorytmów.

Kompozycja pracy jest prawidłowa, a pracę czyta się bardzo dobrze. W mojej ocenie elementy opisujące znane metody, które zamieszczono w rozdziale nr 3 powinny zostać przesunięte do rozdziału nr 2. W przypadku rozdziału nr 4 powinien być on włączony do rozdziału nr 5, gdyż opisuje on *de facto* warunki eksperymentu (miary jakości oraz stosowane narzędzia analizy statystycznej). Redakcja pracy jest na dobrym poziomie, a język rozprawy jest czytelny i poprawny.

Oryginalne osiągnięcia

Do szczególnie oryginalnych i wartościowych elementów rozprawy należy zaliczyć:

- Zaproponowanie znacznego rozwinięcia, zaproponowanej wcześniej w zespole, z którego wywodzi się doktorant, modularnego algorytmu przetwarzania obrazów cyfrowych poprawiającego jakość ich binaryzacji.
- Zaproponowanie kilkunastu wariantów opracowanej metody, poprzez określenie dla kolejnych kroków algorytmu potencjalnych algorytmów, które mogą być w nim zastosowane.
- Zaproponowanie zmodyfikowanej miary oceny jakości dedykowanej segmentacji pisma odręcznego.
- Zaprojektowanie środowiska eksperymentowania oraz ocena zaproponowanej metody na drodze eksperymentu komputerowego z wykorzystaniem zestawu trzech baz benchmarkowych.

Sformułowane algorytmy oraz rezultaty otrzymane w wyniku eksperymentów komputerowych utwierdzają mnie w przekonaniu o ich przydatności w rozważanych przez doktoranta problemach. Cele, sformułowane na początku pracy, zostały osiągnięte, a rezultaty otrzymane na drodze eksperymentów komputerowych zdają się potwierdzać przydatność zaproponowanych algorytmów w rozważanych zadaniach i uprawdopodobniają sformułowaną w rozprawie tezę.

Dorobek publikacyjny autora składa się z 4 publikacji, w tym z artykułu opublikowanego w czasopiśmie indeksowanym w JCR *Applied Artificial Intelligence* (IF=6.725 oraz 200 punktów wg spisu czasopism MEiN). Pozostałe prace zostały opublikowane głównie w materiałach bardzo dobrych konferencji międzynarodowych ICCS, ICVG oraz CISIM (indeksowane w bazie CORE). Przedstawiony dorobek świadczy o zweryfikowaniu przedstawionych w pracy wyników poprzez pozytywny wynik procesu recenzyjnego oraz prezentacje rezultatów rozprawy w trakcie wystąpień konferencyjnych.

Uwagi krytyczne

Lektura rozprawy pozwala sformułować pewne uwagi natury ogólnej oraz szczegółowej. Do uwag ogólnych zaliczam:

- Doktorant wspomniał, że jest autorem algorytmów rozpoznawania tekstów na obrazach, jednakże nie zawarł ich w rozprawie, co nie jest dla mnie zrozumiałe, gdyż jak rozumiem zostały one opracowane w trakcie prac nad dysertacją.

- Co prawda badania eksperymentalne są przeprowadzone dość dokładnie, dla wielu możliwych konfiguracji algorytmu, to moje zastrzeżenia budzi niedokładny opis eksperymentu, tj. nie podano dokładnego protokołu eksperymentalnego. Obecnie w celu weryfikacji badań autorzy udostępniają możliwość wykorzystania opracowanego przez nich oprogramowania. Powyższe warunki są zgodne z podejściem zwanym *repliable research*.
- Nie sformułowano pytań badawczych, na które miały odpowiedzieć wyniki badań esperymentalnych.
- Doktorant przeanalizował wyniki badań eksperymentalnych w wykorzystaniem testu NxN Friedmana oraz post hoc Nemenyi. W pracy brak jest jednak informacji, czy zostały spełnione warunki stosowania tego testu, a także brak jest uzasadnienia jego wyboru (zarówno w zakresie testu NxN, jak i post hoc). Test Friedmana jest znany z tego, że jest czuły na skład metod, które są porównywane. W mojej opinii lepiej stosować testy parowe, jak np. signed-rank test Wicoxona, gdyż w przypadku odrzucenia hipotezy zerowej i tak chcemy otrzymać informacje, w których parach metod możemy obserwować statystycznie istotnie różną jakość.
- W pracy wielokrotnie wskazano na problem efektywności obliczeniowej projektowanych algorytmów przetwarzania obrazów cyfrowych. Jednakże autor nie pokusił się o ocenę złożoności obliczeniowej zaproponowanej metody. Jest to o tyle proste, że dla wielu komponentów zaproponowanego algorytmu złożoność obliczeniowa jest znana.

Do uwag szczegółowych zaliczam:

- Brak umieszczenia spisu skrótów i oznaczeń.
- Wyniki z wykorzystaniem testu statystycznego Friedmana i posthoc powinny zostać przedstawione z wykorzystaniem odpowiednich, powszechnie stosowanych metod wizualizacji.
- Dostrzeżono drobne błędy interpunkcyjne, a także drobne uchybienia redakcyjne, które jednak w żaden sposób nie wpływają na czytelność rozprawy i jednoznacznie pozytywny odbiór pracy.

Przedstawione w recenzji uwagi o charakterze ogólnym mają w większości charakter dyskusyjny oraz uzupełniający i nie wpływają na pozytywne wrażenie o przedłożonej rozprawie, a ich zamieszczenie może być przydatne dla doktoranta w przypadku szerszego publikowania wyników rozprawy, bądź poszukiwania możliwości rozwoju swoich metod.

Konkluzja

Zgodnie z zapisami Ustawy „Prawo o szkolnictwie wyższym i nauce” (Art. 187.1) z dnia 20 lipca 2018r. (dalej:Ustawa) ocenie podlega, czy rozprawa doktorska prezentuje ogólną wiedzę teoretyczną kandydata w dyscyplinie oraz umiejętność samodzielnego prowadzenia pracy naukowej.

Doktorant wykazał się w recenzowanej rozprawie właściwie stosowanym aparatem matematycznym oraz dobrą znajomością aktualnej problematyki związanej z projektowaniem algorytmów na potrzeby wizji komputerowej. Zostało to poparte dobrymi studiami literaturowymi obejmującymi aktualne piśmiennictwo związane z problematyką rozprawy, głównie w zakresie metod poprawy jakości segmentacji. Świadczy to o dobrej wiedzy mgr inż. Kacpra SARNACKIEGO w dyscyplinie Informatyka Techniczna i Telekomunikacja, co jednoznacznie spełnia wymóg Ustawy w tym zakresie.

Recenzowana dysertacja przedstawia rozwiązanie oryginalnego problemu, wzbogacając naszą wiedzę dotyczącą wizji komputerowej, szczególnie w zakresie segmentacji obrazów zawierających pismo odręczne, a zawarte w niej wyniki badań eksperymentalnych wskazują również na możliwość wykorzystania otrzymanych metod w praktyce, co zostało potwierdzone wynikami badań eksperymentalnych. Dla wspomnianych problemów doktorant sformułował szereg ciekawych i użytecznych obserwacji. Mgr inż. Kacper SARNACKI wykazał się zatem umiejętnością samodzielnego prowadzenia badań naukowych, co spełnia wymóg Ustawy w tym zakresie.

Biorąc powyższe pod uwagę stwierdzam, że praca mgr inż. Kacpra SARNACKIEGO pt. *„Poprawa jakości zdjęć pisma ręcznego z wykorzystaniem morfologii adaptacyjnej”* spełnia wymagania stawiane rozprawom doktorskim w świetle Ustawy. Wnoszę o jej przyjęcie i dopuszczenie mgr inż. Kacpra SARNACKIEGO do publicznej obrony.

Michał Woźniak



Częstochowa, 8 stycznia 2022 r.

Prof. dr hab. inż. Krzysztof Cpałka
Katedra Inteligentnych Systemów Informatycznych
Wydział Inżynierii Mechanicznej i Informatyki
Politechnika Częstochowska
Al. Armii Krajowej 36
42-202 Częstochowa

Recenzja rozprawy doktorskiej
mgra inż. Kacpra Sarnackiego
pt. „*Poprawa jakości zdjęć pisma ręcznego
z wykorzystaniem morfologii adaptacyjnej*”,
której promotorem jest
prof. dr hab. inż. Khalid Saeed

1. Zakres tematyczny rozprawy

Recenzja rozprawy doktorskiej została przygotowana na podstawie pisma dra hab. inż. Jarosława Arabasa, prof. Politechniki Warszawskiej, Przewodniczącego Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja. Pismo to jest z 16 listopada 2021 roku.

Recenzowana rozprawa została napisana w języku polskim, składa się ona z 6. rozdziałów, zawierającego 109 pozycji wykazu literatury i streszczenia w językach polskim i angielskim. Całość została przedstawiona na 92. stronach.

Tematyka rozprawy nawiązuje do zagadnienia przetwarzania cyfrowych obrazów dokumentów tekstowych napisanych ręcznie ukierunkowanego na poprawę jakości. Jest to specyficzny obszar przetwarzania obrazów, co wynika z oczekiwań dotyczących wyników takiego przetwarzania. Oczekuje się, że obrazy po przetworzeniu będą obrazami z możliwie najwyraźniej zarysowanymi znakami pisarskimi oraz obrazami biało-czarnymi, aby zminimalizować zajmowaną przez nie przestrzeń dyskową. Jest to szczególnie istotne przy archiwizacji na dużą skalę. Od obrazów liter oczekuje się także uwzględniania typowych dla języka przetwarzanego dokumentu znaków diakrytycznych, a ponadto wyraźnie zarysowanych pętli i ciągłych linii, które w przypadku ich przerwania można próbować na etapie przetwarzania domykać. Tymczasem przetwarzane dokumenty mogą być fizycznie uszkodzone, wyblakłe, nieczytelne, przygotowane na

niejednorodnym kolorystycznie lub ciemnym papierze i napisane po obu stronach kartki, zatem obciążone skutkami przesiąkania tuszu. Takie problemy towarzyszą między innymi przetwarzaniu rękopisów historycznych.

Wymienione kwestie komplikują rozważany w pracy problem i powodują, że zapewnianie dobrej jakości obrazów reprezentujących dokumenty pisane ręcznie jest zagadnieniem złożonym i wieloaspektowym. Można je w praktyce powiązać także z innymi typowymi aspektami przetwarzania obrazów, np. z rozpoznawaniem tekstu, diagnostyką medyczną i przemysłową, analizą podpisów statycznych i dynamicznych, detekcją obiektów na obrazach cyfrowych, analizą zdjęć satelitarnych, kontrolą jakości, biometrią itd. Z tego powodu zagadnienie rozważane w pracy doktorskiej należy uznać za ważne nie tylko z naukowego, ale także praktycznego punktu widzenia.

We *Wstępie* do rozprawy autor w należyty sposób nakreślił rozpatrywany w rozprawie problem, podkreślając jego specyficzny charakter. Na tym tle podał kluczowe cechy opisanego w rozprawie podejścia do polepszania jakości obrazów cyfrowych. Istotne w szczególności wydaje się przytoczenie przez autora artykułu „*Binary handwriting image enhancement by directional field-guided morphology*” opublikowanego w renomowanym czasopiśmie *Information Sciences*, w którym wstępnie wyklarowano założenia opisanego w rozprawie podejścia. Autor rozprawy jest współautorem tego artykułu i we *Wstępie* wskazał on swój wkład w jego powstanie, co nie budzi wątpliwości.

W rozdziale pierwszym rozprawy, zatytułowanym „*Wprowadzenie*” i składającym się z 4. podrozdziałów („*Taksonomia systemów rozpoznawania pisma*”, „*Charakterystyka algorytmów przetwarzania obrazów*”, „*Cel i teza pracy*” i „*Struktura pracy*”), autor dokonał klasyfikacji systemów rozpoznawania pisma wskazując problematykę rozpoznawania pisma odręcznego offline jako jedną z najtrudniejszych. Opisał on także ogólną kilkietapową procedurę rozpoznawania obrazów, od momentu ich akwizycji, przez wstępne przetworzenie (zwane w pracy *preprocessingiem*), segmentację i ekstrakcję cech, po klasyfikację. Ponadto w rozważanym rozdziale sformułowano cel i tezę rozprawy oraz krótko opisano zawartość poszczególnych rozdziałów.

W rozdziale drugim rozprawy, zatytułowanym „*Powiązane prace i stan wiedzy*”, autor dokonał przeglądu literatury pokrewnej tematycznie z tematyką rozprawy. W szczególności szeroko scharakteryzował pojęcie *preprocessingu* obrazu, który może w szczególności zawierać etap *prebinaryzacji*, *binaryzacji* i *postbinaryzacji*. Opisał znaczenie każdego z wymienionych etapów, przytoczył przykłady stosowanych algorytmów, pokazał na przejrzystych przykładach mocne i słabe strony przytoczonych metod. Ważną zaletą przeprowadzonego przeglądu literatury jest także uwzględnienie prac, które nie są bezpośrednio związane z problemem poprawy jakości obrazów

dokumentów pisanych, ale mogą taki proces wspomóc. Przykładowo, rozważano metody identyfikacji naczyń krwionośnych. W mojej opinii rozdział drugi stanowi dobre wprowadzenie w problematykę rozważaną w ramach rozprawy.

W rozdziale trzecim rozprawy, zatytułowanym „*Opis zastosowanych technik*” i składającym się z 6. podrozdziałów („*Zarys zaproponowanej metody*”, „*Binaryzacja*”, „*Morfologia*”, „*Pole kierunkowe*”, „*Test 4-sąsiedztwa*” i „*Zwiększenie liczby próbek*”), autor zaprezentował rozważane w rozprawie podejście do przetwarzania obrazów pisma ręcznego. Jest to podejście mające charakter hybrydowy i można wydzielić w nim następujące po sobie etapy przetwarzania obrazu dokumentu. Wśród tych etapów szczególne miejsce zajmują te, którym poświęcono wymienione wcześniej podrozdziały rozdziału trzeciego. W tych podrozdziałach znalazł się przegląd dedykowanych metod rozważanych w literaturze, który został zrealizowany analogicznie, jak w Rozdziale 2. Podanie wad i zalet przytoczonych metod, w większości użytych następnie do konstrukcji autorskiego podejścia hybrydowego, dobrze uzasadnia strukturę tego podejścia, w szczególności cykl wykonywania użytych metod składowych.

W rozdziale czwartym rozprawy, zatytułowanym „*Weryfikacja przedstawionych technik*” i składającym się z 4. podrozdziałów („*Miara dokładności (Acc)*”, „*Ulepszona miara dokładności (Acc2)*”, „*Test Friedmana*” i „*Test Nemena*”), autor opisał użyte w rozprawie miary oceny skuteczności działania autorskiego podejścia hybrydowego, zarówno w kontekście dokładności wynikającej z porównywania z obrazami wzorcowymi, jak i w kontekście przeprowadzanej analizy statystycznej. Autor rozważał miary odpowiadające przytoczonym tytułom podrozdziałów.

W rozdziale piątym rozprawy, zatytułowanym „*Testy i wyniki*” i składającym się z 6. podrozdziałów („*Parametry algorytmu*”, „*Eksperymenty I (ICFHR)*”, „*Eksperymenty II (DIBCO, PBOK)*”, „*Porównanie z filtrem Frangi’ego*”, „*Analiza statystyczna uzyskanych rezultatów*” i „*Eksperymenty z binaryzacją opartą na sieciach konwolucyjnych*”), autor przeprowadził testy wizualne, analityczne i statystyczne dla trzech różnych baz testowych: ICFHR, DIBCO i PBOK. Pierwsza jest bazą podpisów odręcznych, druga i trzecia są bazami obrazów cyfrowych tekstów pisanych zawierającymi również obrazy wzorcowe. Bazy te różnią się specyfiką obrazów, co pozwoliło autorowi na ocenę odmiennych aspektów działania zaproponowanego podejścia hybrydowego. Cenną zaletą przeprowadzonych i opisanych testów jest uwzględnienie w porównaniach popularnych metod przeznaczonych do binaryzacji i przetwarzania wstępnego. Wzięto pod uwagę w szczególności dwie odmiany dylatacji, zamknięcie i otwarcie morfologiczne oraz filtr medianowy. Autor podjął także próbę użycia sieci konwolucyjnych i filtra Frangi’ego, stosowanego do wymienionego wcześniej problemu identyfikacji struktury naczyń krwionośnych.

W rozdziale szóstym rozprawy, zatytułowanym „Podsumowanie i wnioski”, autor podsumował najważniejsze wnioski wynikające z przeprowadzonych badań oraz nakreślił potencjalne możliwości zastosowania zaproponowanego podejścia w innych obszarach zastosowań, odmiennych od rozważanego w rozprawie.

Pracę kończy sekcja „Bibliografia”, zawierająca wykaz wykorzystanych pozycji literatury. Trzy spośród tych publikacji są autorstwa Pana mgr inż. Kacpra Sarnackiego: dwie zostały opublikowane w materiałach konferencji międzynarodowych (w roku 2018. i 2019.), chronologicznie najnowsza zaś w renomowanym czasopiśmie *Information Sciences* (w roku 2021.), o czym pisano wcześniej.

2. Oryginalne rezultaty uzyskane w rozprawie

Pan mgr inż. Kacper Sarnacki w swojej rozprawie doktorskiej rozważał autorskie podejście do poprawy jakości obrazów rastrowych pisma ręcznego z wykorzystaniem morfologii adaptacyjnej. Zostało ono wstępnie zaproponowane w artykule opublikowanym w roku 2021. w renomowanym czasopiśmie *Information Sciences* („Marcin Adamski, **Kacper Sarnacki**, Kbalid Saeed, *Binary handwriting image enhancement by directional field-guided morphology*, *Information Sciences*, vol. 551, str. 168-183, 2021”). Podejście to charakteryzuje się następującymi cechami:

- Ma strukturę hybrydową i działa etapami. Jest elastyczne w konfiguracji, co skutkuje możliwością dostosowania podzbioru etapów oraz użytych w nich metod nie tylko do oczekiwanej jakości obrazów wynikowych i szybkości ich przetwarzania, ale także do jakości i specyfiki obrazów wejściowych.
- Może współpracować z różnymi podejściami do binaryzacji obrazu, w szczególności metodami: Otsu, Sauvola i Niblacka. Pierwsza maksymalizuje wariancję między dwiema klasami pikseli, może działać na bazie pojedynczego progu lub korzystać z partycjonowania progu, cechuje ją ponadto tendencja do pogrubiania linii, ale przez to również do domykania pętli liter. Druga i trzecia metoda korzystają z wyznaczanego niezależnie dla każdego piksela obrazu progu lokalnego. Różnią się one od siebie m.in. sposobem wyznaczania progu oraz właściwościami: pierwsza jest podatna na utratę ciągłości zapisu znaków pisarskich (glifów) oraz przerywanie pętli, druga dobrze sprawdza się w utrzymywaniu ciągłości zapisu, bywa natomiast podatna na szum. Objawia się to traktowaniem niektórych artefaktów jako litery lub ich składowe.
- Korzysta z morfologii obrazów, w szczególności dwóch typowych podejść: dylatacji i erozji. W praktyce pierwsze z nich zwiększa obiekty na obrazie wpływając na ich obrys, drugie eliminuje z obrazu małe struktury i wąskie linie. Obydwa podejścia korzystają oczywiście

z maski obrazu realizując operację minimum lub maksimum w ramach przetwarzanych nią grup pikseli.

- Korzysta z pola kierunkowego w celu przewidywania kierunku linii służącego korekcie struktur liniowych przerwanych na etapie binaryzacji. Stosuje przy tym trzy podejścia: na bazie gradientu i miary koherencji; na bazie gradientu, miary koherencji i histogramu kołowego oraz na bazie hesjanu. Pierwsze korzysta m.in. z liczenia gradientu realizowanego przez spłot segmentów obrazu z określoną maską (np. Prewitta lub Sobela) w celu ignorowania segmentów tła, a ponadto korzysta ze stosownej miary koherencji do oceny wyznaczanych w segmentach kierunków pola. Drugie podejście eliminuje problem nieprawidłowego wyznaczania więcej niż jednego dominującego kierunku pola w segmentach obrazu przez wykorzystanie m.in. odpowiednio konstruowanego histogramu. Trzecie natomiast podejście bazuje a wyznaczaniu wartości własnych macierzy drugich pochodnych Hessego w celu bardziej precyzyjnej analizy fragmentów obrazu bez potrzeby korzystania z miary koherencji oraz histogramu kołowego.
- Używa testu 4-sąsiedztwa w celu eliminacji z przetwarzanych obrazów niewielkich artefaktów.
- Korzysta z interpolacji kubicznej do normalizacji obrazów oraz czasowego zwiększania rozmiaru obrazu, tj. zwiększania realizowanego na czas przetwarzania w celu użycia większej maski, posiadającej dzięki temu większe możliwości w zakresie reprezentowania elementów liniowych.

Krótko podsumowane oryginalne podejście do poprawy jakości zapisanych w postaci cyfrowej obrazów pisma ręcznego jest zatem umiejętnym połączeniem kilku wybranych metod i zostało ono ściśle dostosowane do specyfiki rozważanego problemu, przy czym wykazano się dbałością o typowe dla tego problemu kwestie. To powoduje, że w praktyce prawdopodobnie trudno byłoby przy takim poziomie dbałości o detale automatyzować proces przetwarzania obrazów cyfrowych pisma ręcznego, korzystając np. z możliwości sztucznych sieci konwolucyjnych. Autor rozprawy odniósł się do aspektów ich użycia w jednym z podrozdziałów prezentujących wyniki testów (Podrozdziale 5.6). Jedną z kwestii utrudniających taką automatyzację mogłoby być sformułowanie funkcji oceny uzyskiwanych rozwiązań, odwzorowującej ludzki sposób postrzegania uzyskiwanych wyników przetwarzania.

Pan mgr inż. Kacper Sarnacki przygotowując rozważaną rozprawę doktorską osiągnął ponadto następujące rezultaty:

- Przygotował środowisko testowe pozwalające oceniać i porównywać zaproponowane warianty podejścia do poprawy jakości obrazów rastrowych pisma ręcznego z wykorzystaniem

morfologii adaptacyjnej. Porównywanie to zostało zrealizowane m.in. w oparciu o dwie intuicyjne miary realizujące sekwencyjne porównywanie pikseli obrazu przetworzonego z dostępnym obrazem wzorcowym. Miary te różnią się od siebie podejściem do tła: druga z nich tło takie eliminuje, aby rozmiar obrazu z dominującym zwykle tłem nie determinował uzyskiwanego rezultatu. W porównaniu uwzględniono także testy statystyczne: Friedmana i Nemenyi. Pierwszy użyto do weryfikacji statystycznej różnic w uzyskanych dokładnościach, drugi natomiast do porównania parami rozważanych algorytmów i ich odmian.

- Przetestował zaimplementowane przez siebie warianty podejścia. Szczególnie ciekawe wydają się rezultaty zaprezentowane w Tabeli 2 „*Miary dokładności (Acc i Acc2) dla wybranych metod binaryzacji i algorytmów preprocessingu, w porównaniu do podejścia autorskiego*” oraz w Sekcji 5.5 „*Analiza statystyczna uzyskanych rezultatów*”. Wynika z nich, że rozważane warianty autorskiego podejścia, niezależnie od użytej metody binaryzacji, działają wyraźnie lepiej od następujących rozwiązań: bez przetwarzania wstępnego, z dylatacją na bazie maski typu diament i typu kwadrat, z otwarciem i zamknięciem morfologicznym oraz od filtru medianowego. Potwierdziły to także wyniki przeprowadzonych testów statystycznych. Niezależnie od przedstawionych w Tabeli 2 i Sekcji 5.5 wartości bazujących na użyciu przyjętych miar dokładności, uwzględniających wzorcowe obrazy pochodzące z baz DIBCO i PBOK, w Sekcji 5.2 „*Eksperymenty I (ICFHR)*” autor przedstawił przykładowe rysunki prezentujące wyniki działania poszczególnych metod w kontekście przetwarzania obrazów cyfrowych podpisów odręcznych. Analiza wzrokowa tych wyników także służy do stwierdzenia, że dla umieszczonych w rozprawie przykładów warianty podejścia autorskiego lepiej radzą sobie np. z problemem naprawy przerwanych linii i problemem eliminacji postrzępienia krawędzi.

3. Uwagi dotyczące rozprawy

Przytoczone dalej kwestie nie wpływają na obniżenie jednoznacznie pozytywnej oceny rozważanej rozprawy, mają one charakter dyskusyjny i są następujące:

- Rysunek 21 na s. 41 prezentuje schemat działania proponowanej metody. Trudno jednak na jego podstawie określić jednoznacznie docelową formułę przetwarzania obrazów, np. jedną z tych, które rozważano w testach i wyklarowano w Tabeli 2. Uważam zatem, że Rysunek 21 mógłby stanowić jeszcze bardziej użyteczny przewodnik po autorskim podejściu do poprawy obrazów cyfrowych pisma ręcznego, gdyby został przygotowany w inny sposób. Przykładowo, mógłby pokazywać on poszczególne fazy przetwarzania obrazu w formie sekwencyjnej z jednoznacznie pokazanymi połączeniami kolejnych metod i wyklarowanymi właściwościami danych wejściowych poszczególnych faz oraz oczekiwanymi rezultatami. Mógłby on

alternatywnie przedstawiać proponowane podejście w formie schematu blokowego z konkretnie zdefiniowanymi blokami warunkowymi, wiążącymi charakter danych wejściowych ze specyfiką odpowiednich metod składowych, co w efekcie mogłoby determinować niejako automatycznie finalny zbiór metod przetwarzania rozważanej klasy obrazów.

- Tabela 2 na s. 71 prezentuje wartości miar użytych do oceny sposobu działania wybranych typowych metod przetwarzania obrazów oraz metody autorskiej występującej w różnych konfiguracjach. Wyniki te jednoznacznie potwierdzają skuteczność rozważanego w pracy podejścia i to nie budzi wątpliwości. Sądzę jednak, że interesujące byłoby przedstawienie tych wyników również w rozbiciu na bazy DIBCO i PBOK, nie zaś wyłącznie w postaci zagregowanej. Pozwoliłoby to wówczas ocenić jakość działania rozważanych metod w kontekście przetwarzania grup obrazów o odmiennej specyfice.
- W Rozdziale 5.6 autor rozważał eksperymenty z binaryzacją opartą na sieciach konwolucyjnych. Jak wcześniej pisałem, trudno byłoby automatyzować proces przetwarzania obrazów cyfrowych pisma ręcznego uwzględniając rozbudowany zestaw różnorodnych i specyficznych oczekiwań typowych dla rozważanego w pracy problemu, wychodzących poza proste stosowane miar dokładności. Oceniam jednak pozytywnie fakt podjęcia próby użycia w swojej pracy przez autora np. sztucznych sieci konwolucyjnych. W mojej opinii być może zabrakło przy tym bardziej szczegółowych informacji o tym, jakiej sieci użyto, jakie miała ona parametry, w jakim środowisku programistycznym wykonano testy, jakie wartości miar uzyskano.
- W problemie przetwarzania obrazów cyfrowych pisma ręcznego istotną sprawą jest redukcja liczby procesów digitalizacji materiałów źródłowych z uwagi na ryzyko zmiany fizycznych właściwości tych materiałów, w szczególności papieru. W tym kontekście bezpieczniej byłoby archiwizować obrazy o dużej rozdzielczości. Z drugiej strony rozmiar pozyskiwanych obrazów cyfrowych powinien być minimalizowany w celu oszczędzania przestrzeni dyskowej. Z jakich rozdzielczości roboczych korzystał autor w trakcie wykonywania testów i czy uważa je za optymalne w kontekście rozważanego przez siebie problemu? Nie chodzi o proces skanowania dokumentów, bo ten z racji korzystania z baz skanów nie był przeprowadzany.
- W pracy występują drobne niedociągnięcia stylistyczne typu: „usprawniający jakość” zamiast „zwiększający jakość” (s. 3), „rozrostem zapotrzebowania” zamiast „zwiększeniem zapotrzebowania (s. 7)”, „będące zbyt małe” zamiast „które są zbyt małe” (s. 57), etc. W pracy występują także drobne błędy edytorskie związane m.in. z zastosowaniem wspólnej numeracji stron dla spisu treści i zasadniczej części pracy, nieposortowaniem numerów pozycji literatury

w odnośnikach do literatury, sposobem przygotowania podpisów pod rysunkami (np. Rys. 7) oraz osadzeniem wzorów w tekście (np. wzorów (2), (3)), które powinny być traktowane analogicznie, jak np. wzór (16).

4. Podsumowanie i konkluzja

W podsumowaniu stwierdzam, co następuje:

- Autor rozprawy doktorskiej rozważał podejście do poprawy jakości obrazów cyfrowych pisma ręcznego z wykorzystaniem morfologii adaptacyjnej, które to podejście należycie przetestował, przytaczając odpowiednio sformułowane konkluzje.
- Rozprawa stanowi dobre podsumowanie ogólnej wiedzy teoretycznej mgra inż. Kacpra Sarnackiego w zakresie dyscypliny Informatyka techniczna i telekomunikacja.
- Rozprawa doktorska zawiera szereg oryginalnych i wartościowych rezultatów naukowych, została zredagowana w sposób poprawny, poszczególne wątki zostały w niej przedstawione w sposób kompetentny. Przygotowanie jej potwierdza umiejętność samodzielnego prowadzenia pracy naukowej przez kandydata do stopnia doktora.

W konkluzji stwierdzam, że praca doktorska „*Poprawa jakości zdjęć pisma ręcznego z wykorzystaniem morfologii adaptacyjnej*”, której autorem jest Pan mgr inż. Kacper Sarnacki, spełnia wymagania obowiązującej ustawy o stopniach naukowych i tytule naukowym. Wnoszę o jej przyjęcie i dopuszczenie do publicznej obrony.

Krzysztof Spätke