



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: ipi@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 Informatycznych Politechniki Warszawskiej pod kierunkiem naukowym Pana prof. dr hab. inż. Khalida Saecda. 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, Sauvoli 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 urzyciem 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 dziedzinowych 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.



