

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 obarczone 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 kilkuetapową 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 Nemenyż*”), 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, Khalid 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 Cpatue