

Warszawa, dn. 7 czerwca 2022

dr hab.inż. Marcin Iwanowski, prof.uczelni
Politechnika Warszawska, Wydział Elektryczny
Instytut Sterowania i Elektroniki Przemysłowej

Recenzja rozprawy doktorskiej mgr.inż. Huberta Michalaka pt. „Poprawa efektywności wybranych metod przetwarzania wstępnego obrazów na potrzeby automatycznego rozpoznawania znaków alfanumerycznych”

1 Zawartość rozprawy

Recenzowana rozprawa liczy 197 stron, w tym 141 stron tekstu zasadniczego. Została napisana w języku polskim, składa się z nienumerowanego wprowadzenia, pięciu rozdziałów oraz trzech załączników.

Wprowadzenie zawiera ogólne omówienie obszaru tematycznego pracy oraz wskazanie na najważniejsze, zdaniem Autora, osiągnięcia. Ponadto zawiera tezę, a także cel i zakres pracy.

Pierwszy rozdział pracy zawiera prezentację kluczowych dla tematyki pracy zagadnień podstawowych takich jak: problem segmentacji obrazu, rozpoznawanie znaków alfanumerycznych, kodów jedno- i dwuwymiarowych, tablic rejestracyjnych oraz innych typów obiektów. Szczególny nacisk został położony na szczegółowe omówienie metod binaryzacji obrazu obejmujące dużą część powszechnie stosowanych metod: Otsu, Kapura, Rosina, Niblacka, Sauvola, Wolfa, Bradleya, NICK, SSP, Fenga, Gatosa, Singha i Bataineha. W osobnym podpunkcie omówiono także zarys metod wykorzystujących uczenia maszynowe.

Rozdział drugi jest omówieniem najważniejszych pojęć związanych z weryfikacją skuteczności działania algorytmów binaryzacji. Przedstawiono w nim podstawowe pojęcia związane z oceną skuteczności i wybrane wskaźniki pozwalające na określanie stopnia odrębności uzyskanego wyniku binaryzacji od danych wzorcowych (ground truth).

Rozdział trzeci jest poświęcony dostępnym zbiorom danych wykorzystywanym w badaniach nad binaryzacją obrazów. Oprócz standardowych zbiorów, z których niektóre zostały wykorzystane w badaniach Autora, są tam także opisane dwa zbiory danych przygotowane w ramach pracy. Pierwszym z nich jest zbiór WEZUT, zawierający zdjęcia sztucznie wygenerowanego tekstu przy zróżnicowanym, nierównomiernym oświetleniu. Drugi zbiór danych zawiera zdjęcia tabliczek znamionowych urządzeń i charakteryzuje się dużym zróżnicowaniem w zakresie warunków oświetleniowych, perspektywy, czy rodzajów aparatów wykorzystanych do akwizycji obrazów.

Najważniejszym rozdziałem pracy jest rozdział 4 zawierający opis metod opracowanych przez Autora. Omówiono w nim najpierw metodę blokową w tym jej wariant podstawowy oraz warianty z jednym i dwoma współczynnikami korekcyjnymi, a następnie z dodatkowym progami nieczułości. Kolejne punkty są poświęcone metodom blokowym wielowarstwowym, metodom bazującym na

estymacji tła oraz metodę wykorzystującą entropię lokalną. W ostatnim podpunkcie omówiono metody hybrydowe.

Ostatni rozdział, piąty, stanowi podsumowanie pracy.

Praca zawiera także trzy załączniki. Załącznik A prezentuje przykładowe obrazy poddawane binaryzacji. W załączniku B znajdują się pełen zestaw 40 obrazów wchodzących w skład bazy obrazów tabliczek znamionowych. Załącznik C zawiera przykładowe wyniki binaryzacji.

Na końcu pracy zamieszczono obszerną bibliografię zawierającą 202 pozycje literatury, w tym 11 pozycji autorskich, oraz spisy tabel i rysunków.

2 Ocena pracy

Praca jest poświęcona jednemu z klasycznych zagadnień przetwarzania obrazu – jego binaryzacji w zastosowaniu do rozpoznawania znaków alfanumerycznych. Zagadnienie to, pomimo długiej historii, jest ciągle żywotnym tematem prac badawczych. Jest to umotywowane jednej strony z faktem, iż – pomimo długiej historii badań – nie wszystkie problemy związane ze zróżnicowaniem obrazów poddawanych binaryzacji zostały ostatecznie rozwiązane. Z drugiej strony pojawiają się nowe obszary zastosowań, takie jak binaryzacja obrazów rejestrowanych telefonem komórkowym, z którymi wiążą się specyficzne problemy.

Opiniowana praca prezentuje wyniki prac badawczych zrealizowanych przez Autora w zakresie binaryzacji obrazu tekstu jako elementu toku przetwarzania obrazu wykonywanego przed właściwym etapem właściwej detekcji znaków tekstu. Teza pracy odnosi się do wpływu udoskonaleń istniejących metod przetwarzania wstępnego na poprawę efektywności działania metod rozpoznawania znaków alfanumerycznych. Poprawa jakości jest przy tym rozumiana zarówno w kontekście wskaźników odnoszących się do skuteczności (dokładności) jak też i do efektywności obliczeniowej. Sformułowany na tej podstawie cel główny pracy został przedstawiony jako analiza metod przetwarzania wstępnego z propozycjami ich modyfikacji ukierunkowanych na poprawę jakości. Oprócz tego został także postawiony cel dodatkowy, którym jest propozycja nowych metod pozwalających na poprawę dokładności metod rozpoznawania tekstu. Warto zaznaczyć, że występujące zarówno w sformułowaniu tezy, jak i celu pracy istotne pojęcie przetwarzania wstępnego nie jest w pracy dookreślone. Może być nim przecież zarówno przetworzenie obrazu przed binaryzacją (np. usunięcie tła ale już nie sama binaryzacja) jak też i całość operacji przetwarzania przed wykonaniem właściwego rozpoznawania tekstu (wówczas binaryzacja także wchodzi w jego skład). Jednak z kontekstu pracy i osiągnięć Autora wynika, że w tym przypadku drugi wariant jest tym właściwym.

Oryginalne osiągnięcia Autora można podzielić na trzy części. Pierwsza część obejmuje metody blokowe. Autor zaproponował metodę binaryzacji z lokalnym doбором progów realizowanym w blokach o ustalonym rozmiarze. Propozycja metody obejmuje cztery jej warianty. Wariant podstawowy polega na prostym podziale obrazu na kwadratowe bloki. Wariant ten został zbadany pod kątem optymalnego rozmiaru bloku. Kolejne trzy warianty są modyfikacjami wariantu podstawowego, dzięki którym nastąpiła poprawa działania algorytmu m.in. w blokach o niewielkiej wariancji jasności. W efekcie dalszych prac została opracowana wielowarstwowa metoda blokowa, w której wartość progowa jest ustalana na podstawie kilku warstw, składających się ze wzajemnie względem siebie przesuniętych, bloków. Dzięki takiemu rozwiązaniu, metoda lepiej radzi sobie z nieciągłościami powstającymi na granicach tych bloków. Opracowanym przez Autora metodom blokowym w ich różnych wariantach zostało poświęcone łącznie pięć artykułów współautorskich.

Drugi obszar oryginalnych osiągnięć Autora dotyczy metod przetwarzania wstępnego obrazu przed binaryzacją, prowadzących do eliminacji efektu zmiennej jasności tła obrazu. Metody te koncentrują się na estymacji tła obrazu poprzez klasyczne podejście polegające na usunięciu z niego obiektów o wielkości zbliżonej do wielkości znaków umieszczonych na stronie tekstu, a następnie na odjęciu takiego tła od obrazu początkowego. Dzięki pozbyciu się tła z obrazu, jego – realizowany w kolejnym kroku – proces binaryzacji właściwej jest realizowany na obrazie o histogramie zbliżonym do bimodalnego. W tym zakresie zostały zaproponowane dwa podejścia. Pierwsze z nich estymuje tło poprzez redukcję rozdzielczości obrazu i następującą po niej ponowne skalowanie do rozmiaru pierwotnego. Dzięki takiemu zestawowi operacji kluczowe detale obrazu czyli znaki tekstu drukowanego są usuwane. Drugie podejście wykorzystuje filtr oparty na lokalnej entropii. Badania Autora w zakresie metod eliminacji tła jako przetwarzania wstępnego przed binaryzacją zostały opisane w czterech współautorskich artykułach.

Trzeci, obszar badań Autora to metody hybrydowe. W zaproponowanym podejściu, wykorzystuje się jednocześnie kilka algorytmów binaryzacji, z których każdy generuje potencjalnie inny, różny od pozostałych metod, wynik. Wynik ostateczny powstaje w efekcie głosowania przeprowadzanego niezależnie dla każdego piksela obrazu (warto przy okazji zwrócić uwagę, że tego typu działanie odpowiada działaniu jednowymiarowego filtra medianowego w sytuacji gdy wszystkie obrazy cząstkowe zostaną połączone w pojedynczą strukturę trójwymiarową, a filtrację przeprowadzi się wzdłuż nowopowstałego w ten sposób wymiaru). Metody hybrydowe był tematem kolejnych dwóch współautorskich artykułów Doktoranta.

Powyższe metody opracowane przez Autora zostały pozytywnie zweryfikowane w przeprowadzonych eksperymentach. Dowodzi to dużego doświadczenia i kompetencji Autora w dziedzinie binaryzacji obrazu, czego potwierdzeniem jest łącznie 11 artykułów naukowych poświęconych tej tematyce, w których Doktorant jest każdorazowo wymianiany jako pierwszy autor publikacji. Zaproponowane rozwiązania są interesujące i oryginalne. Były one nie tylko tematem artykułów naukowych, ale także zostały poddane weryfikacji praktycznej w konkursach, których celem było opracowanie najbardziej skutecznych metod binaryzacji obrazu.

Warto także wspomnieć, że w ramach pracy powstały zbiory danych, przeznaczone do badań nad metodami binaryzacji. Zbiory te zostały publicznie udostępnione, dzięki czemu możliwe będzie ich wykorzystanie do dalszych badań, także przez inne zespoły badawcze.

Prace badawcze opisane w pracy zostały przeprowadzone starannie i zgodnie z właściwie przygotowanym planem badawczym. Praca jest także bardzo staranna pod względem edytorskim, z dobrze dobranym materiałem ilustracyjnym. Zawiera ponadto obszerny przegląd literatury z przedmiotowego zakresu.

3 Uwagi krytyczne

Lektura pracy pozwoliła na sformułowanie następujących dyskusyjnych uwag krytycznych:

1. **Sposób prezentacji wyników poszczególnych grup eksperymentów.** W ramach pracy przeprowadzono wiele cennych eksperymentów. Jednak ocena dla różnych grup metod była przeprowadzana zgodnie z różnymi kryteriami. Współczynniki opisujące jakość przetwarzania dla metod blokowych w p.4.1.1.i 4.1.2 liczone dla pikseli obrazu. Jednak już do oceny następnych, zostały użyte metryki edycyjne operujące na znakach tekstu. W pewnej mierze jest to podyktowane innym charakterem zbiorów na których metody były testowane. Jednak

dla czytelności porównania, optymalnym rozwiązaniem byłoby użycie identycznych metryk i metodologii porównawczych dla wszystkich metod.

2. **Ekstrakcja/usuwanie tła na potrzeby binaryzacji** Zagadnienie usuwania tła z obrazu jest dość dobrze znane w dziedzinie przetwarzania obrazów. Istnieje szereg metod opartych na filtrach kontekstowych, zarówno liniowych, jak i nieliniowych (np. medianowych czy morfologicznych). Zastosowanie tych metod do przetwarzania wstępnego w kontekście głównego tematu pracy tj. binaryzacji obrazów tekstu drukowanego, byłoby cennym tematem badawczym, który nie został jednak rozwinięty przez Autora. Warto było to zrobić, także dlatego, że pierwsza z zaproponowanych przez Autora metod, będąca kombinacją redukcji rozmiaru obrazu i jego powiększenia, jest w istocie metodą, która mogłaby zostać zrealizowana (przynajmniej w niektórych wariantach interpolacji) alternatywnie, poprzez zastosowanie filtru dolnoprzepustowego.
3. **Szczegóły metod blokowych** O ile opisy podstawowych metod binaryzacji znajdujące się w rozdziale 1 pracy są kompletne i wyczerpujące, o tyle opisy metod opracowanych przez Autora pozostawiają pewien niedosyt. Szczegóły metody blokowej (p. 4.1). Zgodnie z jaką zasadą jest przeprowadzania binaryzacja w tym przypadku ? W jakimś sensie problem ten jest wyjaśniony w p. 4.1.5, lecz nie wynika z niego, która z metod została użyta do eksperymentów z metodą blokową i jej wariantami. Współczynniki F-measure z tabeli 4.1 czy też inne, z kolejnych tablic, odnoszą się do (jak można wywnioskować z kontekstu) porównywania piksel-do-piksela. Piksele jakich jednak obrazów są w tym przypadku porównywane ? Jak jest wyznaczana średnia ważona podana w ostatnich wierszach tablic ? Z jakimi wagami zostały uwzględnione poszczególne bazy DIBCO ?
4. **Algorytmy rozpoznawania znaków** Wyniki dużej części eksperymentów zrealizowanych w ramach pracy były weryfikowane metrykami znakowymi. Metryki te wymagają wykonania pełnego zestawu operacji przetwarzania, włączając w to algorytm rozpoznawania znaków. Algorytm ten także może jednak generować błędy, które, w przypadku stosowanej odległości edycyjnej Leveshteina, mogą wpłynąć na ostateczną miarę błędu. Aspekt ten nie został jednak poruszony w pracy. Jaki więc może być wpływ specyfiki zastosowanego algorytmu rozpoznawania znaków na całościową skuteczność badanych metod binaryzacji ?
5. **Inne niż tekst rodzajów treści obrazów** W swojej pracy Autor ograniczył się do binaryzacji zdjęć/skanów tekstu. Jest to zgodne z postawieniem problemu zawartym we wprowadzeniu do pracy, dzięki niemu ujęcie tematu i rozwiązanie problemu są spójne. Niewątpliwie jednak, do pełnej oceny metod binaryzacji, wartym rozważenia byłoby przeprowadzenie testów dla obrazów zawierających inne formy graficzne spotykane w rzeczywistych dokumentach takie jak np. linie, tabele, znaki o innym charakterze i wielkości niż litery. Pozwoliłoby to na bardziej dogłębną analizę badanych metod w zastosowaniu do przetwarzania obrazów dokumentów.
6. **Problem koloru w obrazach** Metody opisane w pracy dotyczą binaryzacji obrazów polegającej na określeniu optymalnej wartości progu. To jest podejście typowe dla obrazów w skali szarości. Jednak na ogół, w tym także w bazach obrazów, wykorzystywanych w badaniach, poszczególne obrazy są kolorowe. Czy i – jeśli tak – to w jaki sposób zostały one przed binaryzacją przekształcone na obrazy w skali szarości ? I szersza kwestia: w jakich

przypadkach informacja o kolorze może mieć wpływ na wynik binaryzacji i jak należałoby tę informację uwzględnić ?

7. **Artefakty kompresji stratnej** Obrazy wykorzystywane w pracy charakteryzują się m.in. zmiennym oświetleniem znacząco wpływającym na binaryzację. Taki dobór obrazów jest jak najbardziej słuszny. Istnieje jednak specyficzna i dość powszechna kategoria zniekształceń obrazu powstających wskutek stosowania algorytmów kompresji stratnej. Szczególnie przy wyższych stopniach kompresji, mogą powstawać zniekształcenia, które mogą mieć wpływ na wynik binaryzacji. Uwzględnienie tego typu zniekształceń pozwoliłoby na ocenę przydatności algorytmów w przypadku pogorszenia jakości obrazów spowodowanego kompresją. Miałoby to istotne znaczenie w przypadku przetwarzania zdjęć dokumentów wykonanych smartfonem, które często są zapisywane z wyższymi współczynnikami kompresji.

4 Uwagi redakcyjne

Z uwagi na dużą staranność edytorską, liczba znalezionych niedociągnięć jest niewielka:

- Na str.116, wiersze 8-9 od góry: "elementu morfologicznego". Takie pojęcie nie istnieje. W metodach morfologicznych otoczenie punktu jest opisywane elementem strukturującym (ang. structuring element), niekiedy – nieco mniej poprawnie – nazywane elementem strukturalnym.
- Na str.90 tabela 4.12, jest "number" obrazu zamiast "numer".
- Na str. 103, koniec trzeciego akapitu, jest "wielości" zamiast "wielkości".

5 Podsumowanie

Podsumowując, niezależnie od zawartych w niniejszej opinii uwag krytycznych, które w większości są elementem dyskusji nad pracą oraz sugestią ewentualnych dalszych prac badawczych, uważam, że Autor pracy wykazał swoje wysokie kompetencje w wybranym przez siebie obszarze badawczym. Recenzowana rozprawa zawiera oryginalne rozwiązania problemów naukowych oraz pokazuje umiejętności Autora w zakresie samodzielnego prowadzenia pracy naukowej. **Całościowa ocena rozprawy jest pozytywna.** Uważam, że rozprawa spełnia wymagania stawiane rozprawom doktorskim przez stosowne, aktualnie obowiązujące, akty prawne. Dlatego **wnioskuję o dopuszczenie rozprawy doktorskiej mgr. inż. Huberta Michalaka do publicznej obrony.**

