

Zielona Góra, 20.06.2022r.

Dr hab. inż. Marcin Mrugalski, prof. UZ
Instytut Sterowania i Systemów Informatycznych
Uniwersytet Zielonogórski
M.Mrugalski@issi.uz.zgora.pl

Recenzja rozprawy doktorskiej

Mgra inż. Huberta Michalaka
pod tytułem

„POPRAWA EFEKTYWNOŚCI WYBRANYCH METOD PRZETWARZANIA WSTĘPNEGO OBRAZÓW NA POTRZEBY AUTOMATYCZNEGO ROZPOZNAWANIA ZNAKÓW ALFANUMERYCZNYCH”

opracowana na zlecenie
Senatu Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie

I. Problem naukowy i obszar rozprawy

Problemem naukowym stanowiącym przedmiot omawianej rozprawy doktorskiej jest opracowanie efektywnych metod binaryzacji obrazów zawierających znaki alfanumeryczne w celu ich rozpoznawania. W literaturze można znaleźć dziesiątki metod binaryzacji podzielonych na kilka podstawowych klas tj. metody bazujące na globalnym lub lokalnym progowaniu (ang. Global/Local Threshold – G/LT), metody bazujące na wykrywaniu krawędzi (ang. Edge-based – EB), metody transformacji obrazów (Image Transform – IT), metody warunkowych pól losowych (ang. Conditional Random Field – CRF), metody wektorów nośnych (ang. Support Vector Machine – SVM), metody uczenia maszynowego (ang. Machine Learning – ML), metody konwolucyjnych sieci neuronowych (ang. Fully Convolutional Network – FCN), metody bazujące na uczeniu głębokim (ang. Deep Learning – DL) czy metody bazujące na uczeniu nadzorowanym i nienadzorowanym (ang. Supervised/Unsupervised Tuning – S/UT).

Pomimo szeregu istniejących algorytmów realizujących zadanie binaryzacji, charakteryzują się one niedostateczną efektywnością działania w warunkach występowania trudności

wynikających z częściowego uszkodzenia nośników przetwarzanych informacji oraz różnorodnych szumów i zakłóceń powstających na etapie procesu digitalizacji.

Liczne źródła zakłóceń w procesie digitalizacji tj. zniekształcenia perspektywy, nierównomierne oświetlenie, refleksy świetlne na powierzchniach metalicznych, rozmycie tekstu, lokalne uszkodzenia fotografowanego lub skanowanego nośnika tekstu w postaci przebarwień, ubytków, rdzy czy zanieczyszczeń powodują niską efektywność większości metod binaryzacji stosowanych w rozwiązaniach praktycznych, co powoduje, iż poruszany w rozprawie temat stanowi nadal otwarty problem naukowy. O jego ważności świadczy również rosnąca ilość tradycyjnych nośników informacji podlegających digitalizacji oraz nowe obszary zastosowania ww. metod choćby takie jak rozpoznawanie obrazów i napisów przez systemy wspierające kierowców pojazdów, czy też systemy stosowane w pojazdach autonomicznych.

Obszar naukowy rozprawy koncentruje się na rozwijaniu istniejących i opracowaniu nowych metod binaryzacji stosowanych w rozpoznawaniu znaków alfanumerycznych. W szczególności podjęto próbę opracowania/rozwoju następujących metod binaryzacji: metod blokowych, metod blokowych wielowarstwowych, metod bazujących na estymacji tła, metod wykorzystujących entropię lokalną oraz metod hybrydowych bazujących na mechanizmie głosowania.

Zgodnie z zaprezentowanym w pracy obecnym stanem wiedzy w rozważanym obszarze badawczym dotychczas nie stosowano tego typu rozwiązań. Zastosowanie tego typu narzędzi motywowane jest faktem, iż są one skuteczne w sytuacjach, gdzie występują częściowe uszkodzenia nośników informacji oraz różnorodne zakłócenia i szumy powstających na etapie procesu digitalizacji.

Celem przedstawionych w rozprawie badań było: opracowanie nowych metod binaryzacji zmierzających do poprawy efektywności rozpoznawania znaków alfanumerycznych przy jednoczesnym ograniczaniu ilości obliczeń bezpośrednio przekładających się na czas przetwarzania. Celem dodatkowym rozprawy było opracowanie nowych metod binaryzacji pozwalających na poprawę dokładności rozpoznawania tekstu na obrazach naturalnych tj. tabliczki znamionowe w systemach automatycznego terowania przesyłek czy opakowań produktów itp. w naturalnych warunkach rzeczywistych.

W rozprawie postawiono następującą tezę: „Zaproponowanie nowych, jak również udoskonalanie istniejących metod przetwarzania wstępnego, w szczególności binaryzacji, wpływa na poprawę efektywności działania systemów bazujących na automatycznym rozpoznawaniu znaków alfanumerycznych. Poprawa ta dotyczy zarówno aspektów jakościowych (skuteczniejsze rozpoznawanie tekstu, lepsze odwzorowanie kształtów znaków, dokładniejsze wykrywanie obiektów itp.), jak również czasowych (redukcja czasu przetwarzania danych, prostsze obliczenia i niższa złożoność obliczeniowa)”.

W świetle tak postawionej tezy, do najważniejszych zadań badawczych mających na celu jej potwierdzenie było:

- analiza właściwości znanych w literaturze metod przetwarzania obrazów pod kątem jakości i szybkości ich działania;

- analiza warunków w jakich funkcjonuje proces przetwarzania obrazów ze szczególnym uwzględnieniem potencjalnych źródeł zakłóceń;
- definicja ilościowych wskaźników jakości procesu przetwarzania obrazów stosowanych do oceny efektywności metod binaryzacji;
- analiza istniejących baz obrazów stosowanych w testowaniu efektywności metod binaryzacji;
- opracowanie nowych i rozwinięcie istniejących metod binaryzacji;
- walidacja efektywności opracowanych metod binaryzacji z użyciem baz obrazów testowych.

II. Koncepcja i struktura rozprawy

Przedstawiona do recenzji rozprawa doktorska zawiera 5 rozdziałów, wprowadzenie, wykaz ważniejszych oznaczeń i skrótów, spisy rysunków i tabel oraz bibliografię. Ponadto, rozprawa zawiera przykładowe bazy danych z obrazami, autorską bazę obrazków tabliczek przemysłowych, oraz przykładowe wyniki przetwarzania obrazów. Podstawę pracy stanowią rozdziały od 1-ego do 5-ego, które stanowią część teoretyczną i badawczą rozprawy.

Wprowadzenie zawiera podstawowe informacje związane z tematyką badań, a w szczególności przedstawia problematykę związaną z procesem przetwarzania obrazów mającym na celu rozpoznawanie znaków alfanumerycznych oraz motywację podjęcia badań w zakresie tego tematu przez autora. Ponadto, prezentuje cel, tezę oraz zakres pracy zawierający krótką charakterystykę jej poszczególnych części.

Rozdział 1 zawiera charakterystykę procesu przetwarzania obrazów ze szczególnym uwzględnieniem etapów segmentacji i binaryzacji. W szczególności w niniejszym rozdziale przedstawiono czternaście metod segmentacji obrazów tj. metody Otsu, Kapura, Rosina, Niblacka, Sauvola, Wolfa, Bradleya, NICK, SSP, Fenga, Gatos, Singha, Bataineha, oraz ogólny zarys metod bazujących na uczeniu maszynowym. Ponadto, rozdział zawiera opis problemu rozpoznawania znaków alfanumerycznych oraz pisma odręcznego, rozpoznawania kodów 1D/2D oraz rozpoznawania napisów na tablicach rejestracyjnych pojazdów oraz innych obiektów rzeczywistych. Dodatkowo, niniejszy rozdział zawiera opis oprogramowania umożliwiającego rozpoznawanie znaków alfanumerycznych.

W **rozdziale 2** przedstawiono przykładowe miary weryfikacji efektów działania algorytmów binaryzacji. W szczególności przedstawiono miary: macierz błędów, raporty z klasyfikacji (czułość, swoistość, precyzja, dokładność, F-Measure, indeks Jaccarda), współczynnik Negative Rate Metric (NRM), współczynnik Misclassification Penalty Metric (MPM), współczynniki Signal-to-Noise Ratio (SNR) i Peak Signal-to-Noise Ratio (PSNR), metrykę Distance Reciprocal Distortion Metric (DRD) oraz miarę odległości edycyjnej. Dokonano też krótkiej analizy metod obiektywnej i subiektywnej weryfikacji jakości obrazu.

W **rozdziale 3** przedstawiono kilka wybranych przez autora rozprawy baz danych obrazów, które są stosowane na etapie weryfikacji efektywności nowo opracowanych metod binaryzacji. W szczególności, przedstawiono bazy: DIBCO, Bickley Diary, Nabuco dataset, LiveMemory,

Brno Mobile OCR Dataset, MonkCuperSet (MCS), SmartDoc, autorskie bazy WEZUT OCR Dataset oraz bazę obrazów tabliczek przemysłowych.

Rozdział 4 zawiera szereg autorskich metod binaryzacji mających zastosowanie w przetwarzaniu obrazów zawierających znaki alfanumeryczne. W szczególności, przedstawiono metody blokowe, metody blokowe wielowarstwowe, metody bazujące na estymacji tła, metody wykorzystujące entropię lokalną oraz metod hybrydowych bazujących na mechanizmie głosowania.

Podsumowanie wyników, najważniejsze oryginalne wyniki pracy oraz kierunki dalszych badań przedstawiono w **rozdziale 5**.

Uzupełnienie pracy stanowią spisy rysunków i tabele. **Bibliografia** obejmuje 202 pozycje literaturowe oraz strony internetowe, które w dobrym stopniu odzwierciedlają istniejący stan wiedzy w niniejszym obszarze badawczym. Warto podkreślić, iż przyjęta koncepcja i struktura pracy jest poprawna.

III. Oryginalne osiągnięcia i znaczenie poznawcze

Przedmiotem pracy badawczej było zaproponowanie nowatorskich metod binaryzacji obrazów zawierających znaki alfanumeryczne, efektywnych w warunkach występowania częściowego uszkodzenia przetwarzanych papierowych lub metalowych nośników informacji oraz różnorodnych szumów i zakłóceń powstających na etapie procesu digitalizacji.

Zgodnie z zaprezentowanym w pracy obecnym stanem wiedzy w rozważanym obszarze badawczym istnieje potrzeba rozwijania tego typu metod. Autor zauważył i przeanalizował szereg interesujących aspektów problemu, które były pomijane lub stanowiły marginalny element w innych opracowaniach. Stąd też konieczne okazało się rozwiązanie wielu cząstkowych zadań. Do najważniejszych oryginalnych osiągnięć rozprawy można zaliczyć:

- Opracowanie metod binaryzacji z podziałem obrazu na bloki i przetwarzanie każdego z tych bloków niezależnie od poprzednich. W szczególności, opracowanie wariantów zaproponowanej metody blokowej z uwzględnieniem współczynnika korekcyjnego, metody blokowej ze współczynnikami a i b , metody blokowej z dodatkowym progiem nieczułości.
- Opracowanie metody blokowej polegającej na wykorzystaniu bloków w strukturze wielowarstwowej pozwalającej na uzyskanie poprawy wyników przetwarzania obrazów. Ponadto, opracowanie mechanizmu dynamicznej zmiany liczby zastosowanych warstw bloków w metodzie binaryzacji blokowej w zależności od wybranego kryterium jakości bądź czasu przetwarzania.
- Analizę możliwości użycia metod estymacji tła w przetwarzaniu obrazów. W szczególności opracowanie metody poprawy jakości przetwarzania dzięki zastosowaniu funkcji przeskalowania zwiększającego i zmniejszającego prowadzącą do estymacji tła pozornego oraz jego eliminacji z procesu binaryzacji.
- Wykorzystanie lokalnej entropii informacji zawartych w obrazie do poprawy skuteczności typowych metod binaryzacji.

- Opracowanie metody hybrydowej z wykorzystaniem mechanizmu głosowania polegającej na wykorzystaniu synergii podstawowych metod binaryzacji obrazów.

Dorobek naukowy Autora rozprawy obejmuje 13 publikacji, w tym 3 artykuły w czasopiśmie (przy czym 2 znajdują się w czasopiśmie z listy JCR tj. Entropy – IF 2.524, oraz Sensors – IF 3.576) oraz 10 artykułów w materiałach konferencyjnych. Wybrane publikacje są cytowane na świecie w materiałach konferencyjnych i artykułach z periodyków znajdujących się na JCR.

Indeksy Hirscha oraz liczba cytowań publikacji w dniu 19.06.22r.

	<i>Scopus</i>	<i>Web of Science</i>
H-index	5	3
Liczba publikacji	12	7
Liczba cytowań	61	30

Uwzględniając wymienione osiągnięcia naukowo-badawcze oraz fakt ich częściowego opublikowania w wyżej wymienionych pozycjach, uważam że mgr inż. Hubert Michalak udowodnił tezę i zrealizował cel rozprawy oraz wykazał się wiedzą i umiejętnością samodzielnego rozwiązywania trudnych problemów teoretycznych i praktycznych szeroko rozumianej współczesnej automatyki i robotyki w zakresie przetwarzania obrazów.

IV. Uwagi i komentarze dotyczące rozprawy

Uwagi ogólne:

1. Str. 51. W rozprawie autor stwierdził „W pracy [180] zostało przeanalizowane użycie trójwarstwowej sztucznej sieci neuronowej z algorytmem wstecznej propagacji błędów (ang. backpropagation learning algorithm) do poprawienia skuteczności rozpoznawania pojedynczych znaków na tablicy rejestracyjnej. Uzyskano, co prawda, poprawę liczby rozpoznanych znaków, jednakże znacząco wydłużył się wówczas czas obliczeń, ponadto niezbędne było wcześniejsze nauczenie sieci.”. Autorzy cytowanej pracy uzyskali 97% skuteczność w rozpoznawaniu znaków alfanumerycznych w czasie 58 ms z użyciem procesora Intel core I3, 2.13GHz co umożliwia zastosowanie opracowanej przez nich metody w czasie rzeczywistym. Czy opinia autora rozprawy nie jest nazbyt krytyczna w odniesieniu do faktu, iż czasy przetwarzania zawarte np. w tabeli 4.11 w rozprawie z użyciem klasycznych metod binaryzacji wydają się być w większości przypadków wyższe?
2. Str. 64. Autor użył stwierdzenia “Typowo stosowane są różnorodne współczynniki służące ocenie skuteczności binaryzacji, jej dokładności, różnorodności odczucia percepcyjnego człowieka, czy określeniu zgodności przetworzonego dokumentu z oryginałem.”. W rozdziale 2 zawarto liczne metody i miary weryfikacji algorytmów

- binaryzacji, aczkolwiek wydaje się, że nie zawarto pośród nich metod odnoszących się do „różnorodności odczucia percepcyjnego człowieka”.
3. Str. 73-76. Autor rozprawy przeprowadza cały szereg analiz wpływu rozmiarów bloków stosowanych w metodzie blokowej na różne współczynniki jakościowe. Z wyników przedstawionych w tabelach 4.1-4.4 można wywnioskować, iż w przypadku wskaźnika F-Measure optymalny rozmiar bloku wynosi 48*48 natomiast w przypadku wskaźnika DRD 16*16, wskaźnika PSNR 32*32, a wskaźnika dokładności klasyfikacji 24*24 i 32*32. Na podstawie uzyskanych rezultatów odnosi się wrażenie, iż opracowana metoda może wymagać żmudnej analizy wpływu parametrów na wskaźniki jakościowe co może powodować radykalne wydłużenie czasu koniecznego do uzyskania optymalnych wyników i w konsekwencji zniechęcać do praktycznego wykorzystania opracowanej przez autora metody binaryzacji.
 4. Str. 77. Autor rozprawy zawarł stwierdzenie „Propozycją udoskonalenia proponowanej metody blokowej jest uwzględnienie współczynnika korekcyjnego. Współczynnik ten, określony jako liczba z zakresu [0...1] dokonuje korekty multiplikatywnej otrzymanego wcześniej progu binaryzacji dla danego bloku [110].” nie przedstawiając w rozprawie szczegółów autorskiej metody binaryzacji. Ponadto, czy możliwe jest opracowanie metody automatycznego doboru optymalnej wartości ww. współczynnika?
 5. Str. 89. Autor rozprawy zawarł stwierdzenie „Podejście tego typu można zastosować również w przypadku omawianych w niniejszym rozdziale metod blokowych, gdzie wprowadza się dodatkowy parametr – próg nieczułości [113].” nie przedstawiając w rozprawie sposobu doboru ww. wartości.
 6. Str. 104. Autor rozprawy przedstawił interesującą technikę poprawiającą jakość binaryzacji bazującą na estymacji tła. Metoda ta jednak ma liczne parametry wymagające doboru ich wartości tj. zastosowana metoda interpolacji, współczynnik przeskalowania, próg binaryzacji. Czy możliwe jest opracowanie metodologii umożliwiającej ich adaptacyjny dobór prowadzący do optymalnych wyników jakości binaryzacji? Z analogiczną sytuacją mamy do czynienia w przypadku metody wykorzystującej entropię lokalną (str. 118).
 7. Str. 138. Stwierdzenie „Przedstawione w rozdziale warianty metody hybrydowej stosującej mechanizm głosowania to wybrane najlepsze przykłady, pokazujące, iż metoda ta posiada duży potencjał do przyszłego rozwoju” rodzi pytanie o generalną skuteczność opracowanych metod binaryzacji bazujących na głosowaniu. Ile z analizowanych metod pozwoliło na uzyskanie dobrych rezultatów jakościowych?
 8. Str. 140. Autor rozprawy użył stwierdzenia „Inną znaczącą trudnością, jaka pojawiła się w czasie prowadzonych badań, był bardzo długi czas przetwarzania przy stosowaniu metod bazujących na uczeniu maszynowym, ... , Trudność ta mocno ograniczała możliwości analizy tych metod, gdyż przeliczenie każdego zestawu obrazów testowych zajmowało znacząco więcej czasu. Dodatkowo zrezygnowano z procesu uczenia tych metod, gdyż te typowo trwają wielokrotnie dłużej.”. Powyższe stwierdzenie budzi wątpliwości

odnośnie wyników przedstawionych w rozprawie dla metod bazujących na uczeniu maszynowym.

Uwagi szczególne:

1. Str. 69. W zdaniu „Wszystkie te obrazy zostały utworzone przy pomocy wysokiej jakości cyfrowej lustrzanki...” ww. stwierdzenie ma charakter potoczny i nie powinno być używane w rozprawie doktorskiej w dziedzinie nauk technicznych.
2. Str. 69. W zdaniu „Składa się ona z 40 zdjęć tabliczek znamionowych urządzeń, tabliczek informacyjnych, czy też tabliczek umieszczanych w pojazdach zawierających m.in. numer VIN samochodu. Wszystkie one zostały sfotografowane w różnorodnych warunkach oświetleniowych, różnymi aparatami, z różnymi perspektywami i są one bezpłatnie dostępne w sieci Internet.” brakuje odnośnika do literatury. Ww. baza jest cennym osiągnięciem Autora natomiast czytelnik rozprawy nie ma do niej dostępu.
3. Str. 138. W zdaniu „na przykład dla metody Sauvoli, NICK czy również metody bazującej na uczelni maszynowym część z obrazów jest nieczytelna, co uniemożliwia rozpoznanie części napisów.” powinno być użyte słowo „uczeniu” zamiast „uczelni”.

V. Podsumowanie recenzji

Reasumując, podniesione wyżej uwagi krytyczne i komentarze nie wpływają jednak na pozytywną ocenę oryginalnych i opublikowanych osiągnięć naukowych, zasadniczych wyników zawartych w recenzowanej rozprawie oraz jej ogólną pozytywną ocenę.

W związku z powyższym stwierdzam, że:

- recenzowana rozprawa doktorska Pana mgr inż. Huberta Michalaka spełnia wszystkie wymagania stosownej ustawy o tytułach i stopniach naukowych.
- wnoszę o przyjęcie opiniowanej rozprawy oraz jej dopuszczenie do publicznej obrony.


dr hab. inż. Marcin Mirnański, prof. UZ