



PRZEWODNICZĄCY RADY DYSCYPLINY
Automatyka, Elektronika i Elektrotechnika
Szela
prof. dr hab. inż. Wojciech Szela

Prof. dr hab. inż. Andrzej Czyżewski
Politechnika Gdańska, Wydział ETI
Katedra Systemów Multimedialnych

30. 12. 2021 r.

Opinia o rozprawie doktorskiej mgr inż. **Karola Piniarskiego**

pt.: „Highly efficient night-vision pedestrian detection based on thermal images” (“Wysoce efektywna detekcja pieszych w obrazach termowizyjnych”),
przygotowanej w Pol. Poznańskiej pod kier. **prof. dr hab. inż. Adama Dąbrowskiego**

Zawartość rozprawy, metodyka rozwiązywania problemów i teza rozprawy

Tematyka pracy przedstawionej na 122 stronach (w tym 48 tablic, 57 rys., 105 poz. bibl.) sytuuje ją w dziedzinie metod przetwarzania obrazu. Autor postawił sobie jako cel udowodnienie tezy, że zbudowane przez niego algorytmy przyczynią się do usprawnienia metody detekcji pieszych przede wszystkim w obrazach termowizyjnych, choć w treści rozprawy poruszane są również zagadnienia przetwarzania obrazów wielospektralnych. Głównym zastosowaniem rozwijanych metod są problemy praktyczne związane z detekcją obiektów w warunkach słabego oświetlenia, bądź w ciemności. Autor sugeruje możliwość użycia usprawnionej przez niego metodyki przetwarzania obrazów przykładowo dla potrzeb sterowania pojazdami autonomicznymi, przy czym zwraca uwagę na ograniczenie wymaganej mocy obliczeniowej, tzn. unika algorytmów, do stosowania których wymagane są zaawansowane procesory graficzne. Postawiony w ten sposób problem jest konsekwentnie rozwiązywany w kolejnych rozdziałach pracy, poprzedzonych analizą stanu rozwoju badań i techniki, przeprowadzoną na podstawie literatury przedmiotu. Treść tego rozdziału erudycyjnego odzwierciedla głęboką wiedzę Autora na temat będący jego treścią i bardzo dobrą znajomość literatury przedmiotu.

Zagadnienie rozwiązywane w pracy należy do aktualnych ze względu na istniejące potrzeby związane z zastosowaniami, ma też wyraźne i jasno określone aspekty naukowe.

Sposób rozwiązania problemu, użyte metody, wymagane umiejętności

Autor postanowił użyć klasycznej metody z przetwarzania obrazu, polegającej na jego segmentacji opartej na progowaniu (thresholding segmentation), aby w szybki sposób ekstrahować niewielką ilość dokładnych regionów zainteresowania (ROI), przy czym założył, że uda się opracować algorytm dokładny i szybki. Przyświeca temu dość proste założenie, że piesi są „cieplejsi”, niż ich otoczenie w nocy. Założenie to w praktycznych warunkach nie zawsze jest jednak słuszne, bowiem podczas

segmentacji pojawia się wiele problemów, które Autor dostrzega i wymienia, takich jak: nierównomierny poziom obserwowanej temperatury tego samego pieszego oraz chwilowa utrata kontrastu termicznego pomiędzy pieszym a otoczeniem. Zagadnienia te były rozważane w referacie opublikowanym z udziałem Autora na konferencji IEEE SPA w 2020 r. ("Improved pedestrian detection by adjustment of segmented ROI in thermal night vision").

Założenie, że sylwetka pieszego charakteryzuje się zwiększonym kontrastem nie zawsze jest potwierdzane w praktyce, co wynika także z analiz, np. widocznych na rys. 7, zamieszczonym w rozprawie. Wniosek wyciągnięty przez Autora po dostrzeżeniu tych trudności jest taki, że wymienione i podobne do nich problemy powinny być kompensowane, aby uniknąć sytuacji, w której pieszy nie jest nie jest wykrywany na etapie segmentacji. W związku z tym Autor stawia pytanie, czy dokładna i stabilna segmentacja pieszych w obrazach termowizyjnych jest możliwa poprzez progowanie. Autor stara się odpowiedzieć pozytywnie na tak postawione pytanie, opracowując i badając metody optymalizacji regionów zainteresowania i wprowadzając mechanizmy ekstrakcji cech dystynktywnych oraz oryginalny mechanizm decyzyjny, oparty na zdefiniowanym przez siebie indeksie (nazwanym „Universal Performance Index”), traktowanym jako miara skuteczności algorytmu, biorąca pod uwagę dokładność i szybkość działania. Aspekt ten został opisany w artykule opublikowanym z udziałem Autora w czasopiśmie Sensors w 2020 r. (pt.: "Tuning of Classifiers to Speed-Up Detection of Pedestrians in Infrared Images").

Jakkolwiek tego typu podejście w epoce zdominowanej przez zastosowania metod uczenia maszynowego wydaje się w dużym stopniu tradycyjne, to Autor broni tego podejścia, wskazując na istotną różnicę w wymaganej mocy obliczeniowej: niskiej w przypadku podejścia klasycznego, znacznie wyższej w przypadku zastosowania bardziej złożonych i nowocześniejszych metod. Z argumentacją dotyczącą mocy obliczeniowej należy się zgodzić, aczkolwiek warto zauważyć szybkie zmiany na rynku procesorów graficznych, ich integrację z popularnymi rdzeniami CPU, upowszechnianie się technologii tanich procesorów graficznych (Intel, Broadcom, ARM, VideoCore). Istnieją też współczesne algorytmy sieci neuronowych, które działają w pełni na CPU, oparte np. na kompresji architektur, wymagające stosunkowo niskiej mocy obliczeniowej. Jednak na korzyść klasycznego podejścia, stanowiącego ośnowę pracy, przemawia analityczny charakter takiego podejścia, wpływający dodatnio na czytelność stosowanych metod i na powtarzalny charakter prowadzonych eksperymentów, co jest zaletą rozprawy z punktu widzenia jej charakteru naukowego i rygoru naukowego, odzwierciedlonych w toku licznych eksperymentów.

Pewien niedosyt w powyższym kontekście budzi to, że czytelnik pracy chciałby zobaczyć porównanie, z którego wynikałoby, na ile dobrze działa w porównaniu z podejściem Doktoranta jakkolwiek współczesny algorytm detekcji pieszych/obiektów (np. algorytm oparty na Faster RCNN), aby ocenić, jak odnosi się dokładność zaproponowanego algorytmu, do algorytmu działającego na procesorze

graficznym (GPU). W takim przypadku lepiej uzasadnione zostałyby zawarte w rozprawie stwierdzenie, że "dzięki uniknięciu potrzeby stosowania procesorów graficznych (GPU) dokładność wzrosła lub zmalała albo pozostała na takim samym poziomie". Z tego też powodu kiedy Autor pisze, że jego wyniki są „w pełni akceptowalne” ("fully acceptable", str. 77) albo „bardzo dobre” (90% to jest rezultat "very good" dla detekcji pieszych – str. 81), to trudno określić na jakiej konkretnie podstawie Autor tak wnioskuje, ponieważ pominięcie 10% pieszych wykrytych np. w autonomicznym pojeździe byłoby problematyczne.

Autor, co prawda, przeprowadził porównanie z klasyfikatorem wykorzystującym sieć CNN (rozdział 6), ale to jednak ograniczony przypadek użycia sieci neuronowej, inny niż wykorzystanie całościowego modelu detekcji pieszych (takiego, jak np. Faster R-CNN). Należy też zauważyć, że sieć CNN, którą stosuje Autor, to AlexNet - czyli historycznie pierwsza współczesna architektura CNN, opracowana w 2012 roku.

Zaproponowana przez Autora metoda ma na celu określenie kilku parametrów. Autor sprawdza ich wpływ w odpowiednich podrozdziałach (np. 3.8.2). Jednak nie jest jasne, czy przeprowadza ocenę w oparciu o zbiór treningowy/walidacyjny/testowy (zwraca uwagę brak informacji w tym zakresie). Jeśli w oparciu o zbiór testowy, to byłoby to podejście wątpliwe od strony formalnej, bo Autor dobiera parametry swojej metody w oparciu o przetwarzanie obrazów ze zbioru, który powinien wykorzystywać dopiero w końcowej fazie eksperymentów. Wobec braku informacji na ten temat recenzent będzie prosił o wyjaśnienie mu tej kwestii, np. w trakcie obrony. Jest oczywiste, że należałoby tego dokonać na zbiorze walidacyjnym, wybrać najlepszy wariant i sprawdzić jak działa na zbiorze testowym.

Dobór parametrów ma znaczny wpływ na wynik końcowy. Recenzent chciałby się więc dowiedzieć z treści rozprawy, w jaki sposób parametry wybrane na jednym zbiorze danych, działają na innym zbiorze danych. Tego rodzaju przypadek użycia ("use-case") mógłby w większym stopniu odpowiadać realnemu zastosowaniu.

Uwagi szczegółowe

- str. 3 Autor pisze: "These techniques do not directly use information about thermal contrast (pedestrians are usually brighter than their surroundings at night)." Recenzent wyraża pogląd, że ten argument nie jest właściwy. Jeśli wytrenowana zostaje sieć CNN do wykrywania pieszych w obrazach termowizyjnych, to model na niej oparty będzie uwzględniał kontrast (co prawda na ogół niejawnie, ale i tak potencjalnie w sposób konkurencyjny do metody opartej o progowanie).

- czy wartość 40% (podana na str. 42) ma odniesienie i uzasadnienie w literaturze?. Takie wymaganie jest stosunkowo łagodne w odniesieniu algorytmu. Tymczasem, znana jest wartość 50%, podana w

szeroko znanej publikacji z 2009 r., zatytułowanej: „Pedestrian Detection: A Benchmark” (badania California Institute of Technology i TU Darmstadt).

- zawartość tabeli nr 35 pokazuje, że klasyfikator CNN działa z dokładnością prawie zawsze około 99%, niezależnie od wielkości obrazu wejściowego. Czy jest tak w istocie, że nawet na obrazie o rozmiarze 16x32 piksele algorytm dobrze rozpoznaje pieszych?. Przydatny byłby komentarz ze strony Autora, aby rozwinąć ten interesujący wątek. Pewne informacje na ten temat znajdują się w dość odległej czasowo (wczesnej) publikacji (IEEE SPA, 2014), więc czy od tamtego czasu Autor zgromadził nowsze wyniki, potwierdzające tak wysoki wynik dla bardzo niskich rozdzielczości?

- str 13 „The CVC-14 and KAIST datasets were used in the experiments...” te dwa zbiory danych nie znalazły odzwierciedlenia w Tab. nr 2.

- str. 43 „The results of single thresholding are shown...” – to dość mało przejrzysty fragment. Czytelnik chciałby się dowiedzieć, jaka wartość MISS-rate oznacza "dobry/akceptowalny wynik"?

- str. 69,70 – Roz. 4 zajmuje tylko 2 strony, a dostrajanie algorytmu ROI zostało zapowiadane we wstępie jako jeden z głównych wkładów rozprawy. Pożądane byłoby odzwierciedlenie ścieżki optymalizacji tego algorytmu

- str. 76, Tab. nr 33 - w tej tabeli należałoby podać liczbę cech (number of features) wykorzystywaną dla potrzeb sieci CNN (tzn. jaka była długość wektora cech bezpośrednio przed klasyfikacją)

Mocne i słabsze strony rozprawy:

- rozdział 2.1. Autor orientuje się znakomicie w zagadnieniach termowizji i potrafił te zagadnienia dobrze opisać

- zasadniczo Autor samodzielnie (od podstaw) zbudował kompletny model algorytmiczny systemu detekcji pieszych, który został potwierdzony w działaniu. W związku z powyższym rozprawa jest potencjalnie przydatna do wykorzystania przez specjalistów konstruujących sprzęt do detekcji termowizyjnej

- rozprawa odznacza się staranną redakcją, dostrzeżone błędy redakcyjne są stosunkowo nieliczne, język obcy nie przeszkodził Autorowi z zrozumiałym redagowaniem treści, potknięcia stylistyczne nie utrudniają lektury w istotny sposób

- opisana procedura generowania ROI (rozdział 3) jest zaawansowana i widoczny jest duży wkład pracy Doktoranta w jej opracowanie (np. w podrozdziale 3.5),

- na str. 40, w roz. 3.6 Autor opisuje oryginalną metodę radzenia sobie przez algorytm z grupami pieszych. Zasługuje ona na uwagę jako oryginalny wkład w algorytmikę przetwarzania obrazu

- w pracy udokumentowana jest imponującą liczbą eksperymentów, użyto wielu zbiorów danych, optymalizowano dobór parametrów. Tabela nr 35 na str. 79 zawiera szczególnie interesujące wyniki, dokumentuje znaczną liczbę eksperymentów (zróżnicowane zbiory danych, różne rozdzielczości obrazu). Odzwierciedla to solidną pracę eksperymentalną przeprowadzoną przez Autora rozprawy. Niestety,

prezentacja wyników bywa miejscami niezbyt przejrzysta (zamieszczono wielką liczbę tabel, wykresów bezpośrednio w tekście pracy, zamiast udokumentowania najbardziej znaczących wyników i przeniesienia pozostałej części do załączników).

Dodatkowe elementy oceny:

- w wyniku realizacji rozprawy Autor nabył umiejętności oraz kompetencji na poziomie pozwalającym na zrealizowanie ambitnego planu badawczego, którego realizacja musiała zostać poprzedzona licznymi pracami związanymi z projektowaniem i implementowaniem algorytmów, a zakończona została staranną obróbką i analizą wyników
- publikacje Autora (wraz ze współautorami) są dostępne w materiałach konferencji IEEE Signal Processing and Applications w latach 2014 (rozszerzona wersja tego referatu ukazała się 2015 r. w czasopiśmie CMST, 20 pkt. MEiN), 2015, 2016, 2020. Ponadto, związany z rozprawą artykuł ukazał się w czasopiśmie Sensors (MDPI, 100 pkt. MEiN)
- biorąc udział w opracowaniu współautorskich publikacji, Doktorant w trakcie pracy nad rozprawą wykazał się również kompetencjami społecznymi niezbędnymi do prowadzenia pracy naukowo-badawczej.

Wniosek

Rozprawa pana mgr inż. Karola Piniarskiego została zrealizowana w sposób odzwierciedlający wysokie kwalifikacje jej Autora, znaczny nakład pracy badawczej, implementacyjnej i eksperymentalnej, jak również staranne podejście do przedstawienia wyników eksperymentalnych i sposobu zredagowania tekstu. W mojej opinii osiągnięcia mgr inż. Karola Piniarskiego spełniają wymogi Prawa o Szkolnictwie Wyższym i Nauce, z dnia 20 lipca 2018 r. (Dz. U. 30. 08. 2018 r. Poz. 1668), stawiane kandydatom do stopnia naukowego doktora.

Ze względu na to, że składające się na rozprawę wyniki prac badawczych i eksperymentalnych zostały opublikowane wielokrotnie na międzynarodowej konferencji oraz dwukrotnie w punktowanych czasopismach, sugeruję jej wyróżnienie, o ile opisany powyżej dorobek publikacyjny spełnia wymogi wyróżniania rozpraw doktorskich w dyscyplinie naukowej automatyka, elektronika i elektrotechnika, prowadzonej w Politechnice Poznańskiej.

