

dr hab. inż. Paweł Forczmański, prof. ZUT  
Wydział Informatyki  
Zachodniopomorski Uniwersytet  
Technologiczny w Szczecinie  
*pforczmański@zut.edu.pl*

## RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY DYSCYPLINY INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI CZĘTOCHOWSKIEJ

Autor rozprawy doktorskiej: **mgr inż. Łukasz Karbowski**

Tytuł rozprawy: **System wykrywania sytuacji nietypowych w obrębie pojazdów z wykorzystaniem obrazów oraz map głębi, bazujący na uczeniu głębokim**

Promotor: **dr hab. inż. Mariusz Kubanek, prof. PCz**

Recenzja została sporządzona w związku z powołaniem mnie przez Radę Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Częstochowskiej (Uchwała nr 7/2024/2025 z dnia 14.11.2024) do pełnienia funkcji recenzenta w postępowaniu o nadanie stopnia naukowego doktora nauk inżynieryjno-technicznych w dyscyplinie Informatyka Techniczna i Telekomunikacja panu mgr. inż. Łukaszowi Karbowskiemu.

### 1. Sylwetka Doktoranta

Pan mgr inż. Łukasz Karbowski jest absolwentem studiów pierwszego (2017) i drugiego stopnia (2018) kierunku Informatyka na Politechnice Częstochowskiej. Jest zatrudniony na stanowisku asystenta na Wydziale Informatyki i Sztucznej Inteligencji Politechniki Częstochowskiej. Dorobek naukowy mgr. inż. Łukasza Karbowskiego, wg bazy Scopus, obejmuje 7 prac opublikowanych w latach 2020 - 2023 (trzy w czasopiśmie i cztery w materiałach konferencji międzynarodowych). Wszystkie związane są z tematyką ocenianej rozprawy. Baza Google Scholar raportuje 10 publikacji

(z lat 2020-2023). Współczynnik wpływu h-index dla publikacji w bazie Scopus jest równy 2, natomiast Google Scholar - 3. Są to wskaźniki typowe dla początkowych etapów kariery każdego młodego pracownika nauki.

## 2. Przedmiot oceny

### 2.1. Zakres, cel, teza i charakter rozprawy

Współczesne systemy bezpieczeństwa pojazdów, takie jak systemy wspomagające kierowcę (Advanced Driver Assistance Systems - ADAS), coraz częściej do monitorowania otoczenia wykorzystują kamery pasma widzialnego i dodatkowe sensory. Niniejsza praca doktorska koncentruje się na propozycji systemu, który wykorzystuje obrazy RGB (red-green-blue) oraz mapy głębi w połączeniu z uczeniem głębokim do wykrywania sytuacji nietypowych w obrębie pojazdów. Celem pracy jest stworzenie rozwiązania, które będzie w stanie szybko i skutecznie identyfikować potencjalnie niebezpieczne sytuacje, takie jak wtargnięcie pieszego lub rowerzysty oraz zwierzęcia na jezdnię bezpośrednio przed pojazdem. Dodatkowo, opracowany system ma reagować w momentach, które mogą być bezpośrednim powodem ww. sytuacji (m.in. zbliżanie się osób i zwierząt do pasa ruchu). Umieszczony w pracy cel jest dość oczywisty ale jego forma jest bardzo rozbudowana i przez to brzmienie niepotrzebnie skomplikowane. Teza jest przeformułowaniem celu pracy i również mówi o możliwości wykorzystania zsynchronizowanych w czasie danych obrazowych i przestrzennych do wykrywania nietypowych sytuacji w otoczeniu pojazdu. Zaprezentowana teza jest zrozumiała, choć jej literalne brzmienie nie jest do końca jednoznaczne („Możliwe jest wykorzystanie zsynchronizowanych obrazów oraz map głębi z wykorzystaniem sztucznej inteligencji do pozyskania informacji, jako wspomaganie wykrycia nietypowych sytuacji otoczeniu pojazdu”). Bez znajomości dalszych części pracy nie jest możliwe ustalenie, czy sztuczna inteligencja służy do pozyskiwania lub synchronizacji, czy to wspomaganie wykrycia wymaga wykorzystania sztucznej inteligencji.

Obecne systemy ADAS, choć zaawansowane, często bazują na analizie pojedynczych źródeł danych, takich jak obrazy z kamer RGB lub proste czujniki odległości. Wykorzystanie map głębi, które dostarczają precyzyjnej informacji o położeniu obiektów względem pojazdu, może znacząco poprawić dokładność i niezawodność takiego systemu. Dodatkowo, zastosowanie głębokiego

uczenia pozwala na automatyczne uczenie się systemu na podstawie dużych zbiorów danych, co umożliwia mu adaptację do różnych warunków i sytuacji.

Bez wątpienia, problematyka omawiana w recenzowanej pracy nie jest zagadnieniem nowym i znanych jest wiele innych prac na ten temat. Dodatkowo, istnieje wiele systemów, zarówno doświadczalnych, jak i produkcyjnych, które realizują podobną funkcjonalność. Pewne nowatorstwo opisanego podejścia polega na umiejętnym zastosowaniu wybranych metod sztucznej inteligencji (konkretnie detektorów obiektów bazujących na głębokim uczeniu) i synchronizacji źródeł danych w celu realizacji postawionego zadania w środowisku mobilnym, bez dostępu do znaczących zasobów obliczeniowych.

Praca ma wyraźny charakter eksperymentalny i praktyczny. Potwierdzeniem słuszności przyjętych założeń jest weryfikacja zbadanych metod na autorskich danych testowych.

## 2.2. Układ rozprawy i jej składowe

Przedstawiona do recenzji rozprawa ma zwartą formę, liczy 128 stron, w tym 10 stron poświęconych jest spisowi literatury. Zawiera standardowe elementy, takie jak wprowadzenie, rozdziały teoretyczne, metodologię, wyniki badań, dyskusję i wnioski. Jej struktura jest prawidłowa. Doktorant w skondensowanej formie oraz w logiczny i czytelny sposób zaprezentował wszystkie istotne zagadnienia związane z przeprowadzonymi badaniami.

## 2.3. Analiza źródeł

Bibliografia przedstawiona w autoreferacie zawiera 151 pozycji obejmujących głównie artykuły publikowane w czasopiśmie zagranicznych (m.in. IEEE Sensors, IEEE Transactions on Multimedia, IEEE Access, IEEE Transaction on Pattern Analysis and Machine Intelligence, MDPI Sensors) referaty prezentowane na konferencjach międzynarodowych (m.in. ICIP, CVPR, ICCV, ECCV) oraz rozdziały z monografii publikowane w latach 2000 – 2023, z czego zdecydowana ich większość to publikacje z ostatnich lat (publikacje dotyczące problemu synchronizacji czasowej – głównie pierwsza dekada XXI wieku, natomiast problematyka widzenia komputerowego – lata 20te XXI wieku). Słabo na tym tle wygląda przegląd istniejących rozwiązań praktycznych, poza symbolicznym odniesieniem do jednego z producentów samochodów (Mercedes-Benz)

oferującego funkcjonalność podobną do tej opisywanej w pracy. Autor nie zauważa rozwiązań typu DRIVE AGX firmy NVIDIA z systemem DriveOS, czy też MobileEye, które umożliwiają przetwarzanie w czasie rzeczywistym dużych ilości danych z kamer, radarów i czujników 3D. Brak jest również odniesienia do znanych wyzwań typu DARPA Challenge<sup>1</sup>, które były prekursorem rozwiązań opisywanych w pracy oraz projektów, które były raportowane w literaturze i realizują praktycznie identyczną funkcjonalność<sup>2,3,4,5</sup>.

Oczywiście trudno oczekiwać, że wśród cytowanych prac naukowych znajdą się wszystkie publikacje związane z tematyką poruszaną w rozprawie, w szczególności z widzeniem komputerowym, ekstrakcją cech, uczeniem głębokim oraz metodami synchronizacji czasowej procesów akwizycji danych, jednak brak odniesienia do ww. projektów jest pewnym niedociągnięciem.

Jednakowoż, nawet biorąc pod uwagę niepełny zakres przeanalizowanej literatury, do której odwołuje się Doktorant, można zakładać, że ma On obszerną i aktualną wiedzę dziedzinową.

## 2.4. Metodyka badań

Zaprezentowana w pracy metodyka badań obejmowała wstępną analizę problemu pozyskiwania danych z kamer pracujących w paśmie światła widzialnego i sensorów typu LIDAR, synchronizację procesu akwizycji danych przez niezależne i nie do końca deterministyczne systemy obrazowania oraz wykrywanie i klasyfikację obiektów w polu widzenia sensorów pojazdu poruszającego się po

---

<sup>1</sup> DARPA Challenge, [https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/DARPA/15-F-0059\\_GC\\_2004\\_FINAL\\_RPT\\_7-30-2004.pdf](https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/DARPA/15-F-0059_GC_2004_FINAL_RPT_7-30-2004.pdf) [dostęp 05.02.2025]

<sup>2</sup> Dza-Shiang Hong, Hung-Hao Chen, Pei-Yung Hsiao, Li-Chen Fu, Siang-Min Siao, "CrossFusion net: Deep 3D object detection based on RGB images and point clouds in autonomous driving", Image and Vision Computing, Volume 100 (2020)

<sup>3</sup> RGB LIDAR Fusion, [https://kangkelvin.github.io/16824\\_RGB\\_LIDAR\\_Fusion/](https://kangkelvin.github.io/16824_RGB_LIDAR_Fusion/) [dostęp 05.02.2025]

<sup>4</sup> Jinsol Choi, Minwoo Shin, and Joonki Paik, "Fusion of an RGB camera and LiDAR sensor through a Graph CNN for 3D object detection," Opt. Continuum **2**, 1166-1179 (2023)

<sup>5</sup> Dai, Y., Kim, D., Lee, K. "An Advanced Approach to Object Detection and Tracking in Robotics and Autonomous Vehicles Using YOLOv8 and LiDAR Data Fusion". *Electronics*, **13**, 2250 (2024)

drodze. Weryfikacja postawionej tezy odbyła się na drodze eksperymentalnej ewaluacji skuteczności proponowanego systemu w kontekście wykrywania potencjalnie niebezpiecznych sytuacji związanych z wtargnięciem pieszego (rowerzysty) lub zwierzęcia na pas ruchu.

Jak już wspomniałem, praca składa się z typowych elementów charakterystycznych dla rozpraw doktorskich w dyscyplinie naukowej informatyka techniczna i telekomunikacja. We wprowadzeniu Autor skupił się na procesie pozyskiwania danych obrazowych za pomocą prostych kamer RGB (zarówno uniwersalnych kamerach webcam, jak i kamerach produkowanych dla przemysłu automotive) oraz procesie pozyskiwania informacji o głębi sceny za pomocą sensorów laserowych typu LIDAR (dwa modele o różnej rozdzielczości pozyskiwanego modelu sceny). Następnie omówił zagadnienie synchronizacji procesów akwizycji danych z pracujących równolegle systemów obrazowania. Zwrócił uwagę na problemy techniczne wynikające z opóźnień komunikacyjnych i zaproponował ich rozwiązanie. W dalszej części Doktorant dokonał przeglądu znanych algorytmów wykrywania obiektów na obrazach cyfrowych, dość dokładnie charakteryzując detektory typu R-CNN (i ich warianty) oraz YOLO. Przegląd uzupełnił omówieniem sieci neuronowych pracujących na chmurach punktów (m.in. PointNet i kolejne) oraz integrujących danych obrazowe i przestrzenne (m.in. Frustum PointNet i inne).

W dalszej części dysertacji, Doktorant przedstawił autorskie rozwiązania i modyfikacje znanych metod. W każdym przypadku opisywał bazowe rozwiązanie oraz dokonywał analizy jego skuteczności, a następnie proponował modyfikacje, polegające głównie na dostosowaniu formatu i zawartości zbioru treningowego oraz sposobu przekazywania danych wynikowych do dalszego etapu przetwarzania. Praca zakończona jest opisem kilku scenariuszy opisujących typowe sytuacje drogowe, związane głównie z wtargnięciem pieszego/rowerzysty lub zwierzęcia na jezdnię bezpośrednio przed pojazdem. Doktorant sprawdził skuteczność opracowanego systemu na sekwencjach wideo, które odpowiadają tym scenariuszom.

Weryfikacja skuteczności zaproponowanych metod odbywała się na drodze eksperymentalnej, zgodnej z ustandaryzowaną metodyką badawczą. Autor po przeprowadzeniu eksperymentów numerycznych potwierdził, że system bazujący na połączonej analizie obrazu RGB i mapy głębi

znacznie lepiej radzi sobie w ww. sytuacjach, niż systemy wykorzystujące pojedyncze źródło danych.

## 2.5. Oryginalność rozwiązania postawionego problemu badawczego

Przedstawiona do recenzji praca stoi na dobrym poziomie badawczym i inżynierskim a aktualność problemu, czyli potrzeba stworzenia alternatywnego systemu identyfikacji zagrożeń w ruchu drogowym, stanowi jej dużą zaletę. Autor precyzyjnie zdiagnozował problemy wynikające z trudności w pewnym i szybkim wykrywaniu obiektów przed pojazdem, które mogą wpływać na bezpieczeństwo jego użytkownika. Zaproponował rozwiązanie, które za pomocą znanych i sprawdzonych algorytmów pozyskiwania, integracji, przetwarzania i rozpoznawania danych obrazowych i przestrzennych, pozwoli na automatyzację procesu identyfikacji zagrożeń i ostrzegania kierowcy. Dużą zaletą prowadzonych badań jest ich ukierunkowanie na budowę prostego, szybkiego i wiarygodnego systemu.

Najważniejsze oryginalne osiągnięcia Autora zawarte w pracy to:

- Propozycja algorytmu synchronizacji procesów akwizycji danych, który uwzględnia heterogeniczną budowę systemu obrazowania wraz opracowaniem specjalistycznej platformy mobilnej;
- Budowa autorskiej bazy danych obiektów zainteresowania w polu widzenia kierowcy oraz trening sieci bazujących na architekturze YOLOv7 oraz PointRCNN i PillarPoints do detekcji tych obiektów;
- Modyfikacja sposobu przekazywania informacji wejściowych w sieciach bimodalnych (łączyjących wejście RGB i dane o głębi) w celu zwiększenia skuteczności detekcji w warunkach utrudnionej obserwacji w paśmie widzialnym;
- Propozycja scenariuszy opisujących sytuację przed pojazdem (wtargnięcie/zbliżanie się obiektów zainteresowania, obecność obiektów w określonych miejscach, identyfikacja kontekstu –

przejście dla pieszych, semafor) oraz sposobów ich wykrywania na poziome logicznym/semantycznym;

- Zbadanie opracowanego rozwiązania na danych rzeczywistych pozyskanych za pomocą autorskiej platformy sprzętowej.

Wymienione powyżej elementy wpisują się jednoznacznie i wnoszą wkład w dyscyplinę naukową Informatyka Techniczna i Telekomunikacja. Co ważne, mają one również charakter interdyscyplinarny wynikający z charakteru danych wejściowych i problemów, które adresują (inżynieria lądowa i transport).

## 2.6. Główne wady rozprawy wraz z krytycznymi uwagami szczegółowymi

Jak w praktycznie każdej pracy naukowej, tak i tu można wskazać szereg nieścisłości i słabości. Są one często wynikiem subiektywnego spojrzenia na omawianą tematykę, ale też część z nich wynika z konfrontacji z powszechną praktyką prowadzenia badań naukowych. Poniżej wymienię kilka wybranych uwag o charakterze dyskusyjnym, które mogłyby być przyczynkiem do dyskusji w czasie obrony pracy:

- Rozdział 2 opisujący formę rastrowych obrazów cyfrowych i wybrane operacje przetwarzania wstępnego jest całkowicie niepotrzebny. Nie wnosi niczego ważnego z punktu widzenia realizowanego tematu, szczególnie że nie wiadomo, czy jakiegokolwiek z wymienionych operacji były wykorzystywane w algorytmach opisywanych w dalszej części pracy;
- Opis formatu zapisu danych pozyskiwanych z LIDARu i szczegóły komunikacji na poziomie pakietów TCP/UDP podane w rozdziale 3 nie wnoszą wiele do podejmowanego problemu, szczególnie że system analizuje gotowe, zapisane uprzednio pliki typu pcap/pcd (przetworzone na etapie uczenia do zwykłych plików tekstowych);
- Wydaje się, że opracowanie systemu synchronizacji, który koryguje przesunięcia między klatkami rzędu mikrosekund nie jest potrzebne w sytuacji, gdy kamera RGB pozyskuje

obraz w czasie kilkudziesięciu milisekund (25 fps) a czas reakcji całego systemu to około 65-100 ms (10-15 fps). Prosiłbym o merytoryczne uzasadnienie tej wątpliwości.

- W pracy brak jest szczegółowych informacji o danych testowych, ich objętości (liczba filmów, rzeczywisty czas nagrań) okolicznościach ich akwizycji (warunki oświetleniowe, pora dnia, opady, nasłonecznienie, oślepianie przez nadjeżdżające z naprzeciwka pojazdy), parametrach ruchu (głównie prędkość jazdy) itp.
- Autor w swoich badaniach wybrał detektor YOLOv7, jednak w momencie pisania pracy detektory YOLOv8-10 były już od pewnego czasu dostępne a raportowana latencja YOLOv8 była nawet 10x niższa, co szczególnie predestynuje go do użycia w proponowanym systemie<sup>6</sup>.
- Trening YOLO oraz sieci PointRCNN i PillarPoints, opisane na str. 84 i 92 były zrealizowane na bazie różnych hiperparametrów i pozwoliły na wybranie tych, które gwarantują optymalny sposób działania detektorów. Procesy ten nie są opisane i nie są znane zakresy zmienności tych parametrów;
- Autor nie zaprezentował kompleksowej oceny skuteczności zaproponowanego systemu, gdyż podawanie jedynie dokładności (accuracy) nie jest wystarczającym czynnikiem pokazującym jego praktyczną wartość. Informacja o fałszywych detekcjach jest również ważna jak procent poprawnych detekcji. System, który zbyt często będzie fałszywie ostrzegał użytkownika jest również mało użyteczny, co system, który będzie pomijał poprawne detekcje.

---

<sup>6</sup> Sapkota, Ranjan; Qureshi, Rizwan; Flores-Calero, Marco; Badgujar, Chetan; Nepal, Upesh; Poulouse, Alwin; Zeno, Peter; Vaddevolu, Uday Bhanu Prakash; Yan, Hong; Karkee, Manoj. (2024). YOLO11 to Its Genesis: A Decadal and Comprehensive Review of The You Only Look Once (YOLO) Series (Preprint)



- Autor wielokrotnie wspominał o możliwości realizacji opracowanego algorytmu w środowisku brzegowym (egde computing), podając jednocześnie informację o jego niskiej responsywności (maksymalnie kilkanaście klatek na sekundę). Nie identyfikuje przy tym wąskich gardeł systemu i nie proponuje rozwiązań, poza użyciem bardziej wydajnego systemu obliczeniowego. Może redukcja rozdzielczości przestrzennej kanału RGB pozwoliłaby na uzyskanie większej wydajności? Może uproszczenie architektury lub zmniejszenie szczegółowości opisu obiektów 3D dało by podobny efekt? Jest to wątek, którego podjęcie w pracy doktorskiej jest bardzo istotne;

W kontekście oceny technicznej recenzowanej dysertacji, należy zauważyć, że jest ona w znacznej większości napisana zrozumiałym językiem, jednak Doktorant nie ustrzegł się pewnych błędów językowych. Widać również dużą różnicę w stylach pisarskich zastosowanych w części opisującej proces synchronizacji procesów akwizycji danych w stosunku do części poświęconej metodom widzenia komputerowego. Ta pierwsza wydaje się być wolna od błędów i stoi na wyższym poziomie redakcyjnym.

Wybrane, najbardziej istotne uwagi edytorskie i techniczne w stosunku do pracy przedstawione zostały poniżej:

- Stosowanie kalki językowej "sztuczne sieci [...] dla obrazów", „autorska adaptacja [...] dla detekcji obiektów”, „sztuczne sieci [...] dla map głębi”;
- Błędy odmiany fleksyjnej i błędy interpunkcyjne (zauważalna liczba);
- Niekonsekwencja w nazewnictwie, np. na str. 27. ujednoliciłbym zniekształcenie „beczkowe” i „beczkowate” jako opisujące to samo zjawisko (ang. barrell distortion); na str. 84 „wartość środka dla etykiety x” i „środek osi y obiektu”, mAP tłumaczona jako „mean average precision” i „mean average percentage” (pierwsze poprawne);
- Schemat na rys. 4.3 nie oddaje poprawnie kątów pokrycia kamery RGB i sensora LIDAR opisanego w tekście (chodzi o kąt widzenia kamery RGB);

- Wykres na rys. 5.14 jest prostym rozwinięciem wykresu z rys. 5.12 (uzupełnionym o dodatkowe wyniki) – jest to powtórzenie, które nie wnosi nic do opisu skuteczności detektorów na bazie YOLO. rys 5.12 jest zatem zbędny;
- Często pozycje literaturowe w nawiasach kwadratowych nie są uporządkowane (m.in. str. 28, 29, 30, 34 i in.);
- Na rys. 6.28 wartość współczynnika Recall dla jednej z klas (trudne do określenia, której) wynosi poniżej 0,3 niezależnie od wartości Confidence, co nie ma odwzorowania w tab. 6.1 (wszystkie wartości powyżej 0,7);
- Rys. 6.38c pokazuje połączone dane RGB i LIDAR, w pracy nie opisano jak zostało to zrobione a wydaje się być to kluczowe dla skuteczności działania detektora;
- Tab. 6.5, 6.6 i 6.7 mogły by być połączone aby porównanie było łatwiejsze;
- Mało staranne zredagowanie spisu literatury, który nie uwzględnia miejsca publikacji (głównie w przypadku prac obecnych w serwisie arXiv).

## 2.7. Znaczenie uzyskanych wyników i ich praktyczne wykorzystanie

Koncepcja prezentowanego systemu i otrzymane wyniki, głównie praktyczne, są interesujące i o dużym potencjale aplikacyjnym w obszarze poprawy bezpieczeństwa ruchu drogowego. Zbadane podejścia mogą również przyczynić się do zwiększenia komfortu jazdy poprzez automatyczne modyfikowanie prędkości i charakteru ruchu pojazdów nieautonomicznych. Badania te są również kluczowe dla rozwoju technologii pojazdów autonomicznych, umożliwiając reagowanie na nieprzewidywalne sytuacje drogowe. Zaprezentowane multimodalne podejście, łączące obrazy, mapy głębi oraz uczenie głębokie, może przynieść znacząco lepsze rezultaty niż tradycyjne, unimodalne metody oparte na tradycyjnym uczeniu maszynowym. Warto dodać, że problematyka synchronizacji źródeł danych w opisywanym problemie nie jest elementem, który jest często podejmowany w publikacjach naukowych dotyczących tematyki automotive a ma bardzo duże znaczenie praktyczne.

### 3. Konkluzja

Recenzowana rozprawa stanowi oryginalne rozwiązanie jednoznacznie sformułowanego problemu naukowego. Autor rozprawy mgr inż. Łukasz Karbowski wykazał umiejętność samodzielnego prowadzenia badań naukowych. Wymienione powyżej uwagi ogólne, polemiczne oraz szczegółowe nie mają znaczącego wpływu na pozytywną ocenę rozprawy. Dodatkowy dorobek naukowy, związany z realizowaną dysertacją, świadczy o dobrej świadomości naukowej Doktoranta.

W związku z powyższym uważam, iż przedstawiona mi do recenzji rozprawa doktorska mgr inż. Łukasza Karbowskiego spełnia wymogi stawiane rozprawom doktorskim przedstawione w Ustawie z dnia 10 marca 2023 r. w sprawie ogłoszenia jednolitego tekstu ustawy - Prawo o szkolnictwie wyższym i nauce (Dz.U. 2023 poz. 742), art. 186 i 187 i niniejszym wnoszę o dopuszczenie jej do publicznej obrony.

Paweł Forczmański