

Recenzja Rozprawy Doktorskiej

mgr inż-a Kamila Tomasza Lelowicza

przygotowana dla Rady Dyscypliny Automatyka, Elektronika i Elektrotechnika
w Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie

Tytuł rozprawy: **Modele matematyczne wybranych komponentów wizyjnych branży motoryzacyjnej i ich badanie w celu rozwoju systemów aktywnego bezpieczeństwa**

Dyscyplina naukowa: **Automatyka, Elektronika i Elektrotechnika**

1. Zagadnienie naukowe rozpatrzone w pracy /teza rozprawy/. Charakter rozprawy (teoretyczny, doświadczalny, inny)? Dyscyplina naukowa.

Rozprawa mgr inż. Kamila Lelowicza dotyczy modelowania toru wizyjnego kamer cyfrowych stosowanych w przemyśle motoryzacyjnym do percepcji otoczenia samochodu w zaawansowanych systemach wspierania kierowcy (ang. *Advanced Driver Assistance Systems – ADAS*).

Tematyka rozprawy, sformułowanie problemu badawczego i metodyka proponowanego rozwiązania pozwalają zaliczyć rozprawę do obszarów badawczych **dyscypliny Automatyka, Elektronika i Elektrotechnika**. Autor zasadniczo posługuje się metodami przetwarzania sygnałów analogowych i cyfrowych w celu modelowania toru wizyjnego cyfrowych kamer wizyjnych oraz metodami wstępnego przetwarzania obrazów 2D w celu rekonstrukcji poprawnej geometrii obrazu i usuwania szumów. Kontekst zastosowania wyników badań w systemach automatyki ADAS istotnie wzmacnia przynależność rozprawy do wnioskowanej dyscypliny naukowej Automatyka, Elektronika i Elektrotechnika.

Teza rozprawy, czyli zagadnienie naukowe (zadanie badawcze) rozpatrywane w pracy, została sformułowana dostatecznie jasno, aczkolwiek określa się w niej jedynie aplikacyjną użyteczność modelowania matematycznego bez specyfikacji metodyki naukowej, tzn. wyjawienia tego, jakie mają być własności czy charakter tych modeli, oraz na czym ma ewentualnie polegać nowatorstwo tej rozprawy. Tym niemniej, zakres poruszanych w tezie zagadnień badawczych wymagających weryfikacji na drodze praktycznych eksperymentów, pozwala zaakceptować podaną tezę jako właściwą dla doktoratu, w którym dominuje aspekt doświadczalnej pracy badawczej i który dostarcza nowej wiedzy technicznej. **Teza rozprawy** brzmi następująco:

„Modele matematyczne komponentów wizyjnych systemów ADAS korzystnie wpływają na proces projektowania i testowania ich funkcjonalności, odwzorowują rzeczywiste czujniki z oczekiwaną dokładnością, dostarczają wymagane dane oraz spełniają ograniczenia wynikające z działania w reżimie czasu rzeczywistego.”

Teza oddaje treść rozprawy, ale nie jest zbyt odkrywczą. Trudno bowiem wyobrazić sobie symulator procesu fizycznego, w tym przypadku tworzenia syntetycznego obrazu otoczenia, bez matematycznego modelu. Wprawdzie rozwój metod aproksymacji funkcji przy pomocy głębokich sieci neuronowych pozwala zastępować modele analityczne neuronowymi modelami obliczeń,

jednak nawet wtedy, przynajmniej na etapie walidacji modeli sieciowych, modele analityczne są niewątpliwie pomocne a często niezbędne.

Podsumowując uważam, że powyższa teza jest akceptowalna dla rozprawy doktorskiej o aplikacyjnym charakterze.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł / w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle / świadczący o dostatecznej wiedzy autora. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Praca zawiera 15 stron spisu literatury (strony 119-133), obejmującego 185 pozycji, w tym 9 pozycji autorskich i współautorskich Doktoranta (pozycje [1]-[9]). Za wyjątkiem publikacji nr [5], pozostałe publikacje i 3 pozycje dotyczące wniosków patentowych (w tym jednego przyznanego), pozostają w bezpośrednim związku z rozprawą doktorską. Sam Doktorant podaje, że „*niniejsza rozprawa doktorska opiera się*” na tych pracach. Zostawia to recenzenta w niepewności co do rzeczywistego wkładu Doktoranta i powinno zostać wyjaśnione poza recenzją.

Uwaga 1. Rozprawa doktorska powinna zasadniczo prezentować nową, niepublikowaną wiedzę, chyba że przedmiotem oceny jest cykl publikacji Doktoranta, a w przypadku publikacji współautorskich, wkład Doktoranta jest jasno zadeklarowany przez współautorów. Sprawa wymaga wyjaśnienia przez Komisję Doktorską.

Autor przeprowadził dość obszerną analizę istniejących technik symulacji wirtualnego środowiska i modelowania czujników wspomnianych trzech typów, stosowanych w praktyce przemysłowej. Łącznie odwołuje się przy tym do 51 prac (pozycje bibliografii od [10] do [60]).

Uwaga 2. Modelowanie czujników opisane jest raczej z punktu widzenia użytkownika niż projektanta takich rozwiązań. Autor wylicza po prostu, że dane prace modelują dany czujnik lub, że model stosowany jest do określonego celu, np. usuwania szumów czy detekcji pasów drogowych, ale rzadko wnika głębiej w metodykę stosowaną w cytowanej pracy lub w zalety takiej pracy.

Tym niemniej przegląd zastosowań omawianych technik w przemyśle motoryzacyjnym świadczy o dostatecznej wiedzy autora w tym zakresie. Końcowe podsumowanie i wnioski z przeglądu źródeł wypadają blado i nie stanowią wystarczającego uzasadnienia dla wyborów dokonanych przez Doktoranta.

Uwaga 3. We wnioskach z przeglądu literatury nie określono jasno co Autor zamierza poprawić w stosunku do istniejących metod modelowania czujników.

W rozdziale 3, w oparciu o bogatą literaturę przedmiotu (pozycje od [58] do [90]), przedstawiono techniki akwizycji obrazu za pomocą kamer wizyjnych i wspomniano o literaturze dotyczącej detekcji obiektów w obrazie. Omówienie akwizycji obrazu (punkt 3.2), Autor rozpoczyna od przedstawienia modelu otworkowego kamery i modelowaniu zniekształceń (aberracji) soczewki (w oparciu o pozycje [58], [61] – [66]). W niektórych wzorach występują niejasności, które omawiam w następnym zagadnieniu (nr 3) recenzji. Kolejny punkt 3.3 rozprawy poświęcono obrazowaniu cyfrowemu sceny za pomocą czujników wytwarzanych w technologiach CMOS i CCD (referowane są pozycje [61], [45], [46] i [67]-[78]). Autor podaje typowe, stosowane w akwizycji obrazu, modele filtracji długości fali światła, mozaikowanie kolorów typu Bayera, RCCC i RYYC_y i schematy demozakowania, modelowanie ilości światła (ekspozycji) podczas akwizycji, wyróżnia problematykę szumów występujących w podzespołach i elementach elektronicznych oraz efekt winietowania. Zasadniczo przegląd w punkcie 3.3 wykonany został poprawnie, miejscami wzbogacony został o wzory i ilustracje. Relatywnie krótki punkt 3.4 przedstawia potok przetwarzania sygnału obrazu (ISP) do czego korzysta się z literatury, pozycje [79, 80, 81].

Uwaga 4. W punkcie 3.4 w zasadzie nie ma nowych informacji, poza niezrozumiałym skupieniem się Autora na reprezentacji obrazu w formacie YUV4:2:0 i troską o to, aby właściwie

interpretować strumień takich danych. Problem ten nie ma żadnej kontynuacji w rozprawie i nie wpływa na proponowane rozwiązania.

Punkt 3.5 poświęcony jest potencjalnym zastosowaniom algorytmów analizy obrazów pochodzących z kamer wizyjnych, zainstalowanych w pojazdach. Autor odwołuje się zasadniczo do pozycji literatury od [82] do [90], przeważnie są to publikacje na konferencjach naukowych, często bardzo renomowanych konferencjach. W tym kontekście zabrakło określenia stanu technologii – jakie zastosowanie znajduje się na którym etapie rozwoju technologicznego – jaki jest stan aktualny, co jest już wdrożone w praktyce przemysłu motoryzacyjnego a co znajduje się na etapie testów w warunkach rzeczywistych lub prac rozwojowych.

Uwaga 5. Brak jest przeglądu aktualnego stanu technologii w zakresie zastosowania metod analizy obrazów rzeczywiście wdrożonych i aktualnie rozwijanych w przemyśle motoryzacyjnym - wykonanego np. na podstawie literatury z konferencji technicznych-branżowych, poświęconych teledystrybucji w pojazdach, czy nawet materiałów reklamowych producentów motoryzacyjnych i wyspecjalizowanych w teledystrybucji dostawców takich firm.

W rozdziale 4 analizowane są modele kamery wykorzystywane w symulatorach środowiska pojazdu, określanymi mianem HFSM. Autor odwołuje się do już wcześniej wymienianych prac (np. [44]-[46], [58]) i do pozycji [91]-[93], dotyczących metody „śledzenia promienia” w grafice komputerowej,

W rozdziale 5 wyjaśniane jest pojęcie „generycznego modelu czujnika w systemie ADAS”. Samo pojęcie czujnika dla osób spoza branży motoryzacyjnej może być mylące, gdyż dotyczy ono oprogramowania symulującego wirtualne środowisko otoczenia pojazdu i symulującego akwizycję cech zadanych wirtualnych obiektów przez modele symulowanych czujników pojazdu. W przeglądzie literatury tego tematu, stanowiącym punkt 5.1 rozprawy, Autor odwołuje się do pozycji bibliografii [19, 29, 30, 31, 102]. W dalszej części rozdziału referowane są częściowo prace już wspomniane w poprzednich rozdziałach, a także do publikacji [102-111] dotyczących programów FCOS3D, Centernet i baz danych dla detekcji obiektów w obrazach kamery i danych z lidaru.

Uwaga 6. Pozycje bibliografii od [112] do [185] nie są w ogóle referowane w treści rozprawy.

3. Analiza merytorycznej treści rozprawy. Czy Autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Rozprawa liczy 137 stron w tym 116 stron zasadniczej treści (strony 17-117), podzielonej na 6 rozdziałów i jeden załącznik. Zasadnicza treść uzupełniona jest spisem treści, listą akronimów (2 strony), bibliografią (185 pozycji, ale tylko 111 z nich jest referowanych w rozprawie), spisem rysunków (62 rysunki) i spisem tablic (8 tablic).

Uwaga 7. Zasadniczą wadą edycji pracy jest zbyt mały rozmiar czcionki na większości schematów blokowych i wykresach z wynikami (na rysunkach 2.1, 2.2, 2.5, 2.6, 2.7, 3.1, 4.3, 4.4, 4.21, 5.1, 5.2, 5.6, 5.7, 5.8, 5.9, 5.10, 5.11), oraz zamieszczanie w schematach blokowych tekstów o szarej czcionce na ciemnym tle. Te niedociągnięcia edycyjne sprawiają, że zrozumienie i merytoryczna ocena pracy są poważnie utrudnione.

Rozdział pierwszy zawiera wstęp, niezbędny do zrozumienia celu i tematyki pracy, ich związku z innymi pracami Autora, poznanie struktury pracy – podziału na rozdziały a na koniec wstępnego rozdziału podano tezę pracy. Teza pracy jest wprawdzie powiązana ze strukturą pracy, jednak oczekiwałbym lepszego ukierunkowania czytelnika na to, gdzie w których rozdziałach będą weryfikowane poszczególne elementy tezy pracy.

Uwaga 8. W szczególności nie zostały przybliżone pojęcia zachowania się czujnika (cytat) „na niskim poziomie”, co jest przedmiotem rozdziału 5, ani modelu (cytat) „komponentów wizyjnych na wysokim poziomie”. Z kolei wymienione w tezie „spełnianie ograniczeń czasu rzeczywistego” nie zostało przypisane do żadnego z rozdziałów tej pracy.

Występuje tu trudność Autora w formułowaniu jasnego, przejrzystego i logicznie spójnego przekazu.

Rozdział drugi, zatytułowany (cytat) „*Rola matematycznych modeli czujników w procesie walidacji algorytmów ADAS*” informuje nas o tym, że Autor zamierza prowadzić wirtualną walidację technik głębokich sieci neuronowych stosowanych do detekcji obiektów 3D w scenie i w tym celu w procesie symulacji percepcji sceny przez rzeczywisty czujnik, zamierza posłużyć się modelem stosowanego czujnika, w tym przypadku cyfrowej kamery wizyjnej. Ta początkowa część rozdziału stanowi motywację prac Autora i moim zdaniem powinna być znaleźć się w rozdziale pierwszym. W ten sposób podczas zapoznawania się z tezą pracy nie byłoby żadnych wątpliwości do czego ma służyć modelowanie kamery w procesie rozwijania algorytmów systemu ADAS.

Zasadnicza część rozdziału drugiego poświęcona jest przeglądowi symulatorów wirtualnego środowiska, tworzonych w oparciu o grafikę komputerową, wykorzystywanych przez branżę motoryzacyjną, przedstawieniu technik symulacyjnych MIL, SIL i HIL oraz przeglądowi technik modelowania podsystemów akwizycji stosowanych w systemach ADS (skanerów, lidarów i kamer), grupowanych jako modele niskiej, średniej i wysokiej jakości (LFSM, MFSM i HFSM, odpowiednio). Autor określa podsystemy akwizycji żargonowo mianem „czujniki”, ale poprawnie należałoby tu mówić o urządzeniach opartych o czujniki danego typu. Niestety towarzyszące opisowi rysunki 2.5 i 2.7 są trudno czytelne a miejscami w ogóle nieczytelne.

Autor przeprowadził wystarczająco obszerną analizę istniejących technik symulacji wirtualnego środowiska i modelowania czujników wspomnianych trzech typów, stosowanych w praktyce przemysłowej. Łącznie odwołuje się przy tym do 49 prac (pozycje bibliografii 10-58). Modelowanie czujników opisane jest raczej z punktu widzenia użytkownika niż projektanta takich rozwiązań.

Uwaga 9. Autor wylicza jedynie że dane prace modelują dany czujnik lub, że model stosowany jest do określonego celu, np. usuwania szumów czy detekcji pasów drogowych, ale rzadko wnika głębiej w metodykę stosowaną w cytowanej pracy lub w zalety takiej pracy.

Rozdział 3 zatytułowany „*System wizyjny*” przedstawia techniki:

1. akwizycji kolorowego obrazu 2D za pomocą kamer wizyjnych,
2. przetwarzanie obrazu za pomocą procesora sygnałowego obrazu (w skrócie ISP),
3. i niektóre algorytmy analizy obrazu, a w zasadzie automatyczną detekcję obiektów w obrazie, w szczególności pieszych.

Wzór (3.2.2) dla modelu otworkowego kamery jest wprawdzie matematycznie poprawny, jednak brak jest wyjaśnienia roli zmiennej ‘ m ’, która powiązana jest z *długością ogniskową* f (gdyby wyróżnić parametr f to zmienna $m = f \cdot Z_{punkt}^{(kamera)}$, czyli byłoby jasne, że zależy od długości ogniskowej i odległości punktu sceny od środka soczewki wzdłuż osi Z układu kamery). We wzorze (3.2.2) ogniskowymi nazywane są dwa parametry f_1 i f_2 .

Uwaga 10. Problem widzę w tym, że we wzorze (3.2.2) parametr skośność s też musiałby być wyrażony w jednostkach relacji miary piksela do miary sceny i uwzględniać w sobie wartość ogniskowej. Tymczasem w praktyce jest on bliski zeru (niemianowanemu) i wyraża wartość funkcji trygonometrycznej dla kąta odchylenia układu osi od prostopadłości.

Przyznaję, że dla samej rozprawy nie ma to znaczenia, gdyż w rozdziale 4 Autor rozwiązuje jedynie uproszczony problem kalibracji, w którym zarówno skośność jak i zróżnicowanie ogniskowej w kierunkach X i Y nie występują. Trudny do interpretacji wzór na m pojawia się dopiero w rozdziale 4 w kontekście kwaternionowego modelu macierzy rotacji.

Uwaga 11. Ani w tym miejscu ani też w innym punkcie rozprawy nie wymieniono algorytmów kalibracji parametrów kamery, zarówno tych znanych z literatury naukowej (np. metody Zhanga zaimplementowanej w OpenCV) jak i typowo stosowanych w przemyśle motoryzacyjnym.

W rozdziale 4 Autor eksperymentuje z dwoma modelami korekty dyspersji radialnej i definiuje problem kalibracji parametrów zewnętrznych kamery, ale samych algorytmów rozwiązania powstających tam problemów optymalizacji nie podaje. Zabrakło przeglądu rozwiązań integrujących wszystkie kroki procesu kalibracji i korekty zniekształceń soczewki.

W punkcie odnoszącym się do modeli zniekształceń soczewki, niespójne z tekstem i najwyraźniej zawierające przekłamania notacyjne indeksów dolnych są wzory (3.2.5-6) i (3.2.11-12). Analizę poprawności tych wzorów i ich komentarza zebrałem w uwagach 12-14.

Uwaga 12. W pierwszych wzorach (3.2.5-6) Autor definiuje korektę zniekształcenia radialnego ($[u_u, v_u]$) jako funkcję oryginalnego obrazu $[u_d, v_d]$, ale jednocześnie podaje, że jest to model zniekształcenia a nie korekty. Wbrew pozorom ma to znaczenie, gdyż prowadzi dalej do błędów.

Uwaga 13. W drugich wzorach (3.2.11-12) najwyraźniej występują błędy w indeksacji zmiennych – definiuje się oryginalny punkt zniekształconego obrazu $[u_d, v_d]$ jako funkcję uprzednio skorygowanych zmiennych $[r_u]$ i tych samych oryginalnych (tylko że centrowanych) zmiennych $[u_d, v_d]$ (!).

Uwaga 14. W obydwu przypadkach podano wzory na wyznaczanie skorygowanych obrazów (korekta), co oznacza też, że notacja drugiego wzoru zawiera błędne oznaczenia indeksów.

W tym punkcie, podobnie jak w przypadku parametrów kamery, także zabrakło włączenia modeli korekcji w typowo stosowane algorytmy korekcji aberracji geometrycznych.

Uwaga 15. Dodatkowo jako niepełność opisu należy uznać brak połączenia obu rodzajów korekty w jedną procedurę korekcji – czy połączenie ma charakter addytywny, czy też ma charakter podstawienia wyniku pierwszej korekty (radialnej) jako wejścia dla drugiej (stycznej)?

Opisu *state-of-the-art* metod kalibracji kamery nie znajdziemy również w rozdziale 4, w którym Autor przedstawia swoje własne rozwiązanie kalibracji parametrów zewnętrznych i ogniskowej kamery. W punkcie 3.4 przedstawiono tzw. potok przetwarzania obrazu (ISP), w którym wymienia się też korekcję obrazu, jednak czyni się to w sposób opisowy, ogólny, bez odwoływania się do wcześniej przedstawionych modeli.

Kolejny punkt poświęcono obrazowaniu cyfrowemu sceny za pomocą czujników wytwarzanych w technologiach CMOS i CCD. Wymienione różne stosowane tu techniki filtracji światła i cyfrowego mozaikowania kolorów, omówiono modelowanie szumów w procesie akwizycji, ekspozycję i winietowanie.

Uwaga 16. Zasadniczo przegląd wykonany został poprawnie, został wzbogacony miejscami o wzory i ilustracje, ale ponownie zabrakło dla wielu poruszanych zagadnień (np. rozkwitu, ekspozycji, usuwania szumów, winietowania) przedstawienia praktycznie stosowanych technik rozwiązywania problemów. Np. podano modele ekspozycji (3.3.2-3.3.5) i szumu (3.3.6), ale pozostaje pytanie, jak z nich się korzysta, czyli jak identyfikuje się parametry modeli i steruje automatyką urządzeń.

Uwaga 17. Na stronie 53 pojawia się fraza „białego Gaussa”, która wydaje się być żargonowym określeniem dla „białego szumu”. Jednak w metodologii analizy sygnałów „szum biały” może ale nie musi mieć rozkładu Normalnego. Jak jest w tym przypadku?

We wzorze na winietowanie (3.3.7) występuje długość ogniskowa f , czyli nie wymaga się odrębnego wyróżnienia f_1 i f_2 . Potwierdza to moją uwagę, że w praktyce sensowne jest wyróżnianie jednej wartości ogniskowej f w modelu kamery i traktowanie f_1 i f_2 jako współczynników korekcyjnych.

Wprawdzie się powtarzam, ale dla porządku wspomnę, że kolejny, relatywnie krótki punkt 3.4 przedstawia w bardzo ogólny, jedynie hasłowy sposób potok przetwarzania sygnału obrazu (ISP).

Uwaga 18. Jedynym wyjątkiem poziomu szczegółowości rozważań w punkcie 3.4 jest (niezrozumiałe w kontekście całej prezentacji) skupienie się Autora na formacie obrazu YUV4:2:0

i sformułowanie wniosku o to, aby właściwie interpretować strumień 10-bitowych danych w komputerze o 8-bitowych bajtach. Przecież redukcja 4:2:0 reprezentacji dotyczy liczby pikseli a nie np. 8-bitowej wartości składowej koloru danego piksela.

Punkt 3.5, poświęcony algorytmom analizującym obrazy pochodzące z kamer wizyjnych, zainstalowanych w pojazdach, zawiera wprawdzie dużo nowej i ciekawej informacji, ale nadal na bardzo ogólnym poziomie. Autor wprawdzie odwołuje się do literatury naukowej, ale przedstawia jedynie zastosowanie technik i nie sugeruje żadnych metod stosowanych w wymienianych tu rozwiązaniach.

Uwaga 19. Również w tym przeglądzie, podobnie jak w przeglądzie dokonany w rozdziale 2, nie dowiemy się na jakim etapie rozwoju technologii znajdują się omawiane tu modele i metody analizy obrazów – dla których technik istnieją gotowe, wdrożone technologie, stosowane w przemyśle motoryzacyjnym, które z nich są w trakcie zaawansowanych prób a które są na wczesnym etapie badań.

Ten punkt, podobnie jak cały rozdział trzeci nie został podsumowany przez Autora, nie wskazał on wniosków, które pozwoliłyby mu logicznie przejść do prac własnych i je umotywić.

W rozdziale 4, zatytułowanym „Model HFSM”, Autor przechodzi do eksperymentalnych badań modeli obrazowania stosowanych w symulatorach kamer wizyjnych. Z faktu istnienia rozwiązań do odwzorowywania kamer i tworzenia syntetycznych obrazów w symulatorach stosowanych w przemyśle motoryzacyjnym, Autor wyciąga wnioski, że polem do badań pozostaje w szczególności problematyka modelowania dystorsji kamery i filtracji kolorów do przestrzeni RYYCy. Dlatego też analizie tych zagadnień i przeprowadzeniu odpowiednich eksperymentów, poświęcono obszerne punkty 4.2 (kalibracja kamery z uwzględnieniem modeli radialnej dystorsji) i 4.3 (model kolorów, w szczególności dla filtrów RYYCy).

Jak wynika ze wzoru (4.2.1) optymalizowaną funkcją celu w analizie porównawczej dwóch metod korekcji radialnej obrazu jest kwadrat błędu – czyli różnica pomiędzy wektorem rzeczywistych promieni referencyjnych a skorygowanych promieni. Jest to oczywiście poprawne. Kolejny wzór (4.2.3) zawiera podstawowy błąd, który prowadzi dalej do błędnych wyników i wniosków Autora.

Uwaga 20. We wzorze (4.2.3) pojawia się niespodziewanie po prawej stronie równania symbol r_u , co oznacza wektor skorygowany – czyli referencyjny w procedurze kalibracji. Jest to zapewne błąd, gdyż powinien to być wektor r_d , czyli przed korekcją. W ten sposób Autor tworzy zależność parametru od samego siebie:

$$r_u = r_u \cdot f(r_d, k), \text{ czyli } 1 = f(r_d, k)$$

Metodyka eksperymentu opisana w punkcie 4.2.1 jest dość niezwykła – weryfikacja obu metod korekcji radialnej odbywa się nie na podstawie danych rzeczywistych ale na podstawie innego modelu.

Uwaga 21. Referencyjny model $r_u = f(r_d)$ ma postać krzywej zaznaczonej na rysunku 4.2. Jest to Najwyraźniej występują w nim dwa człony – liniowy i wykładniczy, a krzywa została wygenerowana na podstawie jakiegoś nieznanego wzoru. Dla możliwości powtórzenia eksperymentu przez Czytelnika należy wyjaśnić w jaki sposób powstał ten referencyjny model.

W takiej sytuacji odtworzenie eksperymentu było utrudnione, ale było niezbędne po to, aby zweryfikować wyniki i wnioski podane w punkcie 4.2.1, które z powodu zastosowania przez Autora błędnego wzoru (4.2.3) przypuszczalnie są błędne. Dane odczytane z wykresu na rysunku 4.2 pozwoliły recenzentowi przybliżyć funkcję referencyjnego modelu jako:

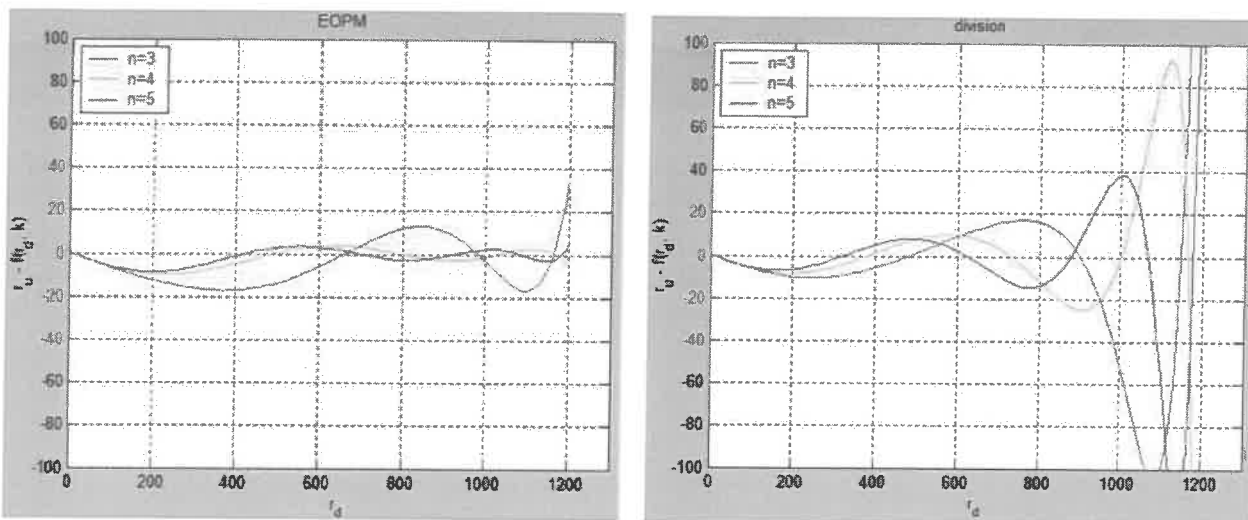
$$r_{est} = 12 * \exp((r - 10)/178 - 1) + 0.9 * r - 3.7$$

Wnioski z odtworzonego eksperymentu zawarłem w kolejnych uwagach 22-23.

Uwaga 22. W wyniku odtworzenia eksperymentu z uwzględnieniem poprawnego wzoru dla (4.2.3) uzyskano wyniki znacząco sprzeczne z wynikami i wnioskami Autora (patrz rys. 1 poniżej). Model EOPM zachowuje się stabilnie w całej badanej dziedzinie, zarówno w jej części

liniowej jak i nieliniowej. Nie obserwujemy na końcu dziedziny znaczących błędów co miał zaobserwować Doktorant. Takie błędy z kolei występują dla modelu „division” i są dużo wyższe na końcu dziedziny niż podane w rozprawie.

Uwaga 23. W eksperymencie Recenzenta model EOPM generuje również mniejsze błędy niż model „division” w całej dziedzinie począwszy już od liczby parametrów $n=4$.



Rys. 1 Porównanie modeli korekcji dystorsji radialnej: EOPM i „division”.

W punkcie 4.2.2 opisano współautorską metodę detekcji narożników szachownicy w obrazie kalibracyjnym z dokładnością subpikselową w oparciu o głęboką neuronową sieć splotową. Autor określa te punkty krawędziowe mianem „punktów siodłowych” w nawiązaniu do interpretacji funkcji jasności obrazu na wykresie 3D. Opisy sieci i algorytmu uczenia sieci są zasadniczo zrozumiałe i poprawne, gdyż stosowane są tu znane techniki.

Uwaga 24. Niejasny jest wzór (4.2.7), gdyż: (1) operuje na zmiennej x , gdy tymczasem ma być stosowany do obu współrzędnych – x , y ; (2) nie wiadomo, czy współrzędne wejściowe są z zakresu 0-1 czy też są nimi wyniki tej funkcji.

Weryfikacją metody są przykładowe obrazy szachownicy, w którym wykryto wszystkie narożniki. Jest to ciekawe rozwiązanie i zapewne obliczeniowo efektywne. Dla lepszej weryfikacji należy jednak porównać ją z istniejącymi algorytmami analizy obrazu szachownicy dostarczającymi punkty narożnikowe pól szachownicy na potrzeby kalibracji (np. dostępnymi w bibliotece OpenCV), pod względem skuteczności i wymagań sprzętowych.

Podobne wrażenie odizolowania od innych metod sprawia algorytm kalibracji opisany w punkcie 4.2.3. Motywacja zastosowania notacji kwaternionowej nie została przekonująco przedstawiona. Ten rozdział oparto o publikację współautorską [6] (Springer, KKA 2020).

Uwaga 25. W istocie ta metodyka jest stosowana dla reprezentacji macierzy rotacji wokół dowolnej prostej w grafice komputerowej, ale nie zostały wyjaśnione jej zalety w kontekście problemu optymalizacji dla kalibracji kamery?

Nie dostrzegam tu zalet stosowania tejże notacji w tym co twierdzi Autor, że (cytat): „Przeformułowanie problemu z wykorzystaniem algebry kwaternionów pozwala na uniknięcie problemu związanego z nakładaniem ograniczeń podczas optymalizacji.”

Uwaga 26. Jaka jest zaleta reprezentacji kwaternionowej, bo sugerowane w rozprawie ograniczenie (ortogonalność macierzy) dotyczy przecież samej macierzy rotacji a nie sposobu zapisu jej parametrów.

Autor kończy omówienie procesu kalibracji parametrów zewnętrznych kamery ogólnym przedstawieniem procedury numeracji punktów narożnikowych pól szachownicy i nie wyjaśnia w jaki sposób te parametry zostaną obliczone (estymowane).

Uwaga 27. W opisie procesu kalibracji brak jest funkcji celu i sposobu rozwiązania problemu – np. czy rozwiązanie ma postać analityczną (np. metodą macierzy pseudo-odwrotnej Moore-Penrose, czy SVD w iteracjach procedury Iterative Closest Point), czy też stosowana jest jakaś metoda gradientowa w problemie nieliniowej optymalizacji?).

Ponownie pojawia się prosty błąd edycyjny we wzorach (uwaga 28).

Uwaga 28. Najwyraźniej równania (4.2.22) i (4.2.23) nie służą do wyznaczania v_u i u_u , jak podaje Autor, tylko do v_u^* i u_u^* , jak wynika z kontekstu. Zresztą wkrótce potem Autor podaje, że v_u i u_u wyliczane są na podstawie (4.2.18) (czyli nie są to równania 4.2.22-4.2.23).

Zastosowanie logarytmu w funkcji celu procedury korekcji radialnej zapewne ma na celu ograniczenie dużych błędów modelu „division” występujących w końcowym zakresie dziedzin promienia. W zasadzie Autor to potwierdza w tekście. Jednak model EOPM w eksperymencie Recenzenta (zakładając poprawność danych) nie wykazuje takich dużych błędów. Stąd wynika kolejna uwaga 29.

Uwaga 29. Uważam, że dodatkowa nieliniowa funkcja logarytmu w funkcji celu metody najmniejszych kwadratów nie byłaby potrzebna, gdyby zastosować poprawnie zdefiniowany podstawowy model korekcji radialnej EOPM.

Punkt 4.3 stanowi niemal pełne tłumaczenie współautorskiego artykułu Doktoranta opublikowanego w „IEEE Sensors Journal” (pozycja [4] bibliografii), w którym przedstawiono modele kolorów RGB i RYYCy, oraz zaproponowano i sprawdzono dokładność dwóch metod konwersji obrazu RGB na surowy obraz RYYCy – powstałych dzięki wytrenowaniu sieci neuronowej typu Resnet lub optymalizacji numerycznej 15 parametrów modelu wielomianowego. Zastosowana w uczeniu sieci neuronowej metoda „perturbacji wag” wydaje się bazować na publikacji współautorskiej [5] (IEEE, MMAR 2021).

Tym razem wyniki autorskich metod porównano ze standardowym modelem konwersji i dwoma dalszymi rozwiązaniami podanymi w publikacjach [100, 101], wykazując poprawność i konkurencyjność własnych metod. Ponownie mam uwagę co do niepełnego opisu warunków przeprowadzonego eksperymentu (uwaga 30).

Uwaga 30. Przedstawiony opis ponownie nie jest pełny, gdyż Autor pominął charakterystykę zbioru danych użytych do trenowania i testowania metod. W rozprawie mowa jest o jedynie o zebraniu par obrazów RGB i RYYCy, ale nie podano na ilu przykładach trenowano sieć neuronową i zidentyfikowano parametry modelu wielomianowego a na ilu testowano wszystkie metody.

Należy podkreślić to, że liczba parametrów modelu wielomianowego (15) stanowi jedynie 1% liczby wag sieci neuronowej a pomimo to średnie błędy wynoszące 1% i 2% różnią się co do wartości niewiele. Tym niemniej należy docenić to, przedstawiono dużą pracę badawczą skutkującą nowymi metodami konwersji przestrzeni barw, co jest niewątpliwym osiągnięciem naukowym Autora i współautora omawianej publikacji (i punktu 4.3 rozprawy).

Niestety w opisie wyników ponownie występuje wiele oczywistych przekłamań notacyjnych. Zebrałem je w kolejnych uwagach 31 – 34.

Uwaga 31. We wzorze (4.3.20) zabrakło podniesienia wyrażenia do kwadratu (miała to być norma L2), a w obu metrykach (4.3.20) i (4.3.21) zabrakło podzielenia wyniku przez N (obie miały wyrażać średni błąd kwadratowy względnie bezwzględny ale nie błąd sumaryczny, jak to wynika z podanych wzorów).

Uwaga 32. We wzorze (4.3.21) nie powinno też być sumowania po trzech kanałach, skoro zastosowano go osobno do każdego kanału (wyniki w tabeli 4.3). Wzór (4.3.20) stosowany jest dla uzyskania wyniku średniego po wszystkich kanałach (tabela 4.2) a wzór (4.3.21) dla średnich dla każdego kanału z osobna (tabela 4.3).

W praktyce czytelnik rozumie, że są to skróty myślowe Autora, ale niepotrzebnie świadczą one o swobodnym operowaniu wzorami w rozprawie zajmującej się „modelami matematycznymi”.

Uwaga 33. Kolejny błąd edycyjny to określenie, że wyniki w tabeli 4.4 uzyskano przy zastosowaniu miary (cytat) „ J_1 (4.3.21)”, gdy numer wzoru odnosi się do miary J_2 .

Uwaga 34. W tabeli 4.4, wyniki dla sześciu metod opisano jako „czerwony-żółty-żółty” zamiast „czerwony-żółty-cyan”.

Kolejne uwagi 35-36 dotyczą stosowanej procedury porównania metod i wskazują na wątpliwość co do poprawności danych liczbowych, zawartych w tabelkach 4.2 i 4.3.

Uwaga 35. Pojawia się wątpliwość co do poprawności zastosowanej metodyki porównywania metod konwersji do RYCy z obiema metodami konwersji do RYB, polegającą na jednostronnej interpolacji wyników RYB do przestrzeni RYCy. Niewątpliwie taki dodatkowy krok transformacji danych wprowadza dodatkowy błąd estymacji składowych co pogarsza wyniki dla tych metod. Aby osiągnąć uczciwe porównanie z tymi metodami należałoby symetrycznie wyznaczyć odwzorowanie RYCy do RYB i także wtedy porównać wyniki tych metod.

Uwaga 36. Porównanie ze sobą wyników podanych w tabelach 4.2 (liczonych według normy L2) i 4.3 (liczonych według normy L1) stawia pod znakiem zapytania ich poprawność. Skoro wyniki dla składowych są praktycznie zawsze wartości ułamkowej (np. dla metody NN wszystkie trzy kanały wykazują średni błąd L1 wynoszący ok. 0.02) to wartość średnia według normy L2 powinna być znacznie mniejsza niż średnia po kanałach liczonych według normy L1 (idealnie powinna wynosić $0.577 = 3^{1/2} / 3$) (np. dla NN według L2 powinna wynosić ok. 0.0115 zamiast podanej w tabeli wartości 0.023). Dla metody podstawowej jest podobnie – średnia według L1 ma wynosić ok. 0.1 a średnia według L2 - 0.103 (a nie rzędu 0.58 jak sugerują wzory), a także dla modelu wielomianowego „Poly” (średnia L1 ok. 0.04, podano średnią L2 0.046 zamiast ok. 0.023).

Rozdział 5 p.t. „Model MFSM” zawiera badania tzw. generatywnego czujnika w systemie ADAS.

Punkt 5.1 omówiłem w przeglądzie literatury. W punkcie 5.2 przedstawiono algorytm wykrywania stopnia przesłaniania obiektu sceny „widzianego” w symulowanym, syntezowanym obrazie, stosując technikę śledzenia promienia, typową dla 3-wymiarowej grafiki komputerowej. Autor stosuje w rozprawie pojęcie „generyczny model czujnika” - należy jednak uściślić, że w przypadku „czujnika” chodzi o „czujnik ADAS”, czyli cały podsystem do akwizycji i przetwarzania danych z rzeczywistych czujników zainstalowanych w pojeździe, a „model generyczny” oznacza symulację rzeczywistej akwizycji cech obiektów wirtualnego środowiska. Algorytm został opatentowany (poz. [7]), a jego opis oparto o publikacje współautorską [8] (Springer, KKA 2020).

Informacją wejściową dla tego algorytmu Autora jest symboliczny opis środowiska, zawierający m.in. położenie, orientację i prędkość obiektów znajdujących się w otoczeniu kamery pojazdu.

Algorytm stanowi uproszczenie metody „śledzenia promieni” stosowanej w grafice komputerowej do tworzenia syntetycznego obrazu 2D na podstawie opisu powierzchni 3D obiektów sceny, z możliwością cieniowania kolorów ścianek przy zadanym oświetleniu sceny. Proponowany w rozprawie algorytm określa jedynie liczbę przecięć obiektów wirtualnej sceny przez promienie przechodzące od środka optycznego przez zadaną siatkę punktów obrazu do obiektów sceny. Na podstawie relacji liczby widocznych i niewidocznych przecięć oblicza się stopień widoczności obiektu. Algorytm zaimplementowano jako osobny program testowy i wykonano badania zależności błędu detekcji od gęstości siatki obrazu oraz zależności czas wykonania programu od liczby wątków (stopnia równoleglenia wykonania).

Uwaga 37. Na marginesie: równanie (5.2.2) to prawo Amdahla a nie Amhdala, jak błędnie podano w rozprawie.

Przejdźcie do kolejnego punktu „5.2.3 Parametryzacja modelu” Autor motywuje następująco (cytat): „*W celu lepszego odzwierciedlenia rzeczywistych algorytmów model ten został*

sparametryzowany. Sposób oraz przykład parametryzacji modelu GSM został przedstawiony w artykule [9]. W istocie punkt 5.2.3 jest tłumaczeniem rozdziałów 4-8 artykułu współautorskiego [9] i obejmuje testowanie dwóch programów FCOS3D (poz. [13]) i CenterPoints (poz. [12]) o otwartym kodzie na zbiorze Nuscene (poz. [108]) oraz ich zastosowanie do estymacji funkcji prawdopodobieństwa detekcji pojazdów i pieszych, sparametryzowanej wynikami autorskiej metody określania rozmiaru widocznego obszaru obiektu i stopnia jego widoczności. Jest to bardzo ciekawa i obszerna praca eksperymentalna, która posłużyła Autorom do eksperymentów z przejściami dla pieszych w celu określania prawdopodobieństwa detekcji pieszego przesłanianego przez jeden samochód przez oba badane programy detekcji analizujące dane z sensora (kamery wizyjnej lub lidar) „naszego” pojazdu, w zależności od odległości od przejścia i prędkości pojazdu.

Muszę jednak zauważyć, że ponownie zabrakło informacji o sposobie rozwiązania problemu optymalizacji i wynikach. W artykule [9] może nie było na to wystarczająco dużo miejsca, ale w rozprawie można było podać dokładniejszą weryfikację metody, co zawarłem w uwadze 38.

Uwaga 38. Ponownie opis problemu optymalizacji kończy się na sformułowaniu ogólnej postaci funkcji prawdopodobieństwa (bez wyjaśnienia powodów wyboru jej postaci i liczby parametrów) a sposób rozwiązania problemu optymalizacji ani uzyskane wartości parametrów nie zostały podane. Uniemożliwia to zainteresowanemu czytelnikowi powtórzenie eksperymentu Autora i weryfikację wiarygodności prezentowanych w rozprawie wyników.

Podsumowując analizę treści rozprawy stwierdzam, iż poszczególne elementy rozprawy są ze sobą spójne, proponowane rozwiązania są oparte o nietrywialne modele, aczkolwiek rozwiązania te stanowią tłumaczenie kilku, przeważnie współautorskich, publikacji Doktoranta. Jednocześnie obserwuję szereg niedostatków, nie tylko o edycyjnym ale także merytorycznym charakterze, które wymagają wyjaśnienia przez Autora. Kilka proponowanych rozwiązań jest oryginalnych (nowe metody konwersji do przestrzeni barw RYYCy, wyznaczanie prawdopodobieństwa detekcji obiektu przez program analizujące dane z czujnika w zależności od rozmiaru obiektu i stopnia jego widoczności) i odpowiada poziomowi techniki reprezentowanemu przez literaturę światową.

4. Słabe strony rozprawy i jej główne wady.

Praca zawiera dużą liczbę wad polegających na błędnych wzorach, niepełnych opisach rozwiązań i warunków przeprowadzania eksperymentu oraz szereg niedostatków edycyjnych:

1. Proponowana metoda korekcji radialnej jest najwyraźniej błędna a wyniki Doktoranta nie mogą być potwierdzone eksperymentem wykonanym przez Recenzenta.
2. Wyniki eksperymentu konwersji kolorów budzą wątpliwości i wydają się być niezgodne ze stosowanymi miarami błędu.
3. Opisy metod korekcji i kalibracji kamery zawierają jedynie informacje o modelach i częściowo funkcje celu, ale nie zawierają informacji o sposobach rozwiązania wymaganych w nich problemów optymalizacji.
4. Opisy warunków prowadzenia eksperymentów w rozdziale 4 są niepełne. Nie podano liczebności zbiorów danych, środowisk implementacji metod oraz stosowanych bibliotek programowych.
5. Przegląd literatury odnośnie metod kalibracji i korekcji zniekształceń kamery nie zawiera opisu metod wcześniej proponowanych w literaturze. Nie porównano osiągniętych wyników z wynikami innych metod znanych z literatury.

Szczegółowe uwagi zawarte są w punkcie 3 recenzji i wyróżnione nagłówkami „Uwaga 1, 2, ..., 38”.

Zwracam też uwagę Komisji Doktorskiej, że rozprawa zawiera znacząco dużo treści, która była uprzednio publikowana, częściowo w publikacjach współautorskich. Podstawowe osiągnięcia w rozdziałach 4 i 5 są tłumaczeniami znaczących części publikacji współautorskich. Również wnioski patentowe, na które powołuje się Autor, są współautorskie. Uważam, że w takiej sytuacji Komisja powinna posiadać deklaracje współautorów tych publikacji o wkładzie Doktoranta w tych publikacjach.

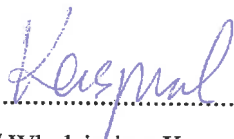
5. Jaka jest przydatność rozprawy dla nauk technicznych i/lub jej innowacyjność technologiczna?

Z uwagi na uprzednio wskazane wady edycyjne pracy (częściowo nieczytelne ilustracje), niepełne opisy metod i warunków wykonania eksperymentów, a także błędy w podstawowych wzorach, jej przydatność dla nauk technicznych jest przeciętna. Natomiast można potwierdzić innowacyjność pewnych proponowanych technik (metody transformacji koloru do przestrzeni RYYCy, algorytm detekcji stopnia i obszaru zasłaniania obiektu w symulacji obrazu kamery, określanie prawdopodobieństwa detekcji pieszego częściowo przesłanianego przez samochód przez oprogramowanie do analizy obrazu względnie danych z lidar).

6. Końcowa ocena

Rozprawa posiada wartościowe elementy, jej treść jest spójna a Doktorant posiada dobry dorobek w zakresie prac badawczych. Z uwagi na wskazane w recenzji błędy, wątpliwość co do części prezentowanych wyników, niepełne opisy metod i warunków przeprowadzenia eksperymentów oraz niedostatki edycyjne, rozprawę należałoby skierować do poprawy, jednak nie jest to formalnie możliwe.

Końcowa ocena rozprawy Kamila Tomasz Lelowicza jest **pozytywna**. Zaliczam rozprawę do kategorii „**spełniająca wymagania**” i wnoszę o dopuszczenie jej do dalszych etapów przewodu doktorskiego.


.....
/ Włodzimierz Kasprzak /