



**Politechnika
Śląska**

WYDZIAŁ AUTOMATYKI, ELEKTRONIKI I INFORMATYKI

KATEDRA AUTOMATYKI I ROBOTYKI

**Statystyczne podejście do weryfikacji i walidacji
systemów sterowania w pojazdach autonomicznych**

Praca doktorska

Autor: mgr Paweł Kowalczyk

Promotor: prof. dr hab. inż. Jacek Izydorec

Opiekun przemysłowy: dr inż. Marcin Szelest

Gliwice, wrzesień 2022

Spis treści

1	Wprowadzenie	1
1.1	Droga do autonomii	1
1.1.1	Percepcja samochodu	2
1.1.2	Aktywne systemy bezpieczeństwa	4
1.1.3	Gromadzenie danych	4
1.1.4	Testowanie i analiza jakości percepcji	5
1.2	Problem badawczy	6
1.2.1	Motywacja	6
1.2.2	Tezy rozprawy	7
1.2.3	Granice zakresu rozważań	8
1.2.4	Zagadnienia omawiane w rozdziałach	9
2	Budowa i testowanie inteligentnej percepcji pojazdu	11
2.1	Pojazd autonomiczny	11
2.1.1	Definicja	11
2.1.2	Klasy autonomiczności	13
2.1.3	Pojazdy autonomiczne dostępne na rynku	14
2.2	Samochodowe „zmysły”	15
2.2.1	Kamery	15
2.2.2	Radary	17

2.2.3	Lidary	18
2.2.4	Czujniki ultradźwiękowe i mikrofony	19
2.3	Moduły percepcji	20
2.3.1	Definicja	20
2.3.2	Dane referencyjne	20
2.3.3	Moduł wykrywający pieszych	21
2.3.4	Pojazdy na drodze	24
2.3.5	Znaki i sygnalizacja świetlna	24
2.3.6	Linie, droga i wolna przestrzeń	24
2.4	Testowanie percepcji pojazdu - wyzwania	24
2.4.1	Testowanie niedeterministyczne - dystrybucja danych	26
2.4.2	Przechowywanie i dostęp do danych	26
2.4.3	Etykietowanie danych	27
2.4.4	Resymulacja	28
3	Ewaluacja interpretacji otoczenia	31
3.1	Prostokątne regiony rozpoznania	31
3.1.1	Inne klasy danych	32
3.2	Wyznaczanie jakości modułu percepcji	33
3.2.1	Ocena jakości detekcji	34
3.2.2	Indeks Jaccarda	34
3.2.3	Miary jakości detekcji oraz ich zastosowania	35
3.2.4	Detekcja prawdziwie pozytywna	37
3.2.5	Podsumowanie jakości w czasie	38
3.2.6	Wpływ dokładności detekcji na układ sterowania	38
3.3	Wybór reprezentatywnego zbioru danych	39
3.3.1	Charakterystyka dużych zbiorów danych	40

3.3.2	Wizualizacja złożonych danych	41
3.3.3	Analiza trajektorii	42
3.3.4	Metryka Wassersteina	43
4	Nowa miara podobieństwa prostokątnych regionów rozpoznania	47
4.1	Podobieństwo prostokątnych obszarów rozpoznania	47
4.1.1	Podobieństwo pola	48
4.1.2	Podobieństwo kształtu	48
4.1.3	Podobieństwo pozycji	49
4.1.4	Generalna miara podobieństwa	52
4.1.5	Dopasowanie detekcji między źródłami	53
4.2	Analiza zdarzeń w czasie	55
4.2.1	Miara podobieństwa sekwencji	56
4.2.2	Średnia ważona dla późnych detekcji	57
4.2.3	Rola pierwszej detekcji	60
4.3	Wizualizacja jakości sceny testowej	63
4.3.1	Rozpoznania fałszywie pozytywne	64
4.4	Kalibracja miar jakości dopasowania	66
4.4.1	Wymagania kalibracyjne wybranych modułów	66
4.4.2	Analiza danych radarowych i lidarowych	73
4.5	Porównanie miar jakości lokalnej	75
4.5.1	Porównanie własności miar	75
4.5.2	Eksperyment Monte Carlo	78
5	Meta-analiza jakości dla zbiorów scenariuszy testowych	83
5.1	Dane i opis eksperymentu	83
5.1.1	Tworzenie układu eksperymentalnego	84
5.1.2	Algorytm klastrowania jakościowego scen	85

5.1.3	Rodzaje analizowanych zaburzeń	90
5.2	Wyniki	93
5.2.1	ANOVA w podgrupach	93
5.2.2	Analiza topologiczna	100
6	Charakterystyka dużych zbiorów trajektorii	111
6.1	Rekonstrukcja trajektorii ze zbiorów detekcji	111
6.1.1	Siatka zajętości	112
6.1.2	Podobieństwo trajektorii	113
6.2	Analiza scenariuszy testowych	116
6.2.1	Mapa termiczna zajętości otoczenia	116
6.2.2	Porównywanie scenariuszy testowych	118
6.3	Porównanie zbiorów scenariuszy	121
6.3.1	Kumulatywne mapy termiczne	122
6.3.2	Klastrowanie rozkładów zajętości	123
6.3.3	Współczynniki różnorodności	124
7	Podsumowanie	127
7.1	Konkluzja o zrealizowaniu celu pracy	127
7.2	Dalsze badania	132
	Bibliografia	150

Spis rysunków

1.1	Diagram ilustrujący uproszczony schemat działania samochodu wyposażonego w inteligentną percepcję	2
1.2	Diagram ilustrujący proces testowania modułów percepcji od momentu zebrania danych testowych, ich etykietowania do ewaluacji wyników uzyskanych przez resymulację	5
2.1	Przykłady produkowanych seryjnie modeli samochodów o niezerowym poziomie autonomiczności	15
2.2	Kamera przednia wkomponowana w lusterko wsteczne samochodu marki Audi [34]	16
2.3	Przykładowe modele radarów marki Continental przeznaczone dla samochodów autonomicznych. Radar krótkiego zasięgu po lewej [38] i radar długiego zasięgu po prawej [39]	17
2.4	Lidar zamontowany na dachu samochodu testowego na torze Aptiv [42] . .	18
2.5	Czujnik ultradźwiękowy [44] (po lewej) i system mikrofonów [45] (po prawej)	19
2.6	Przykładowe dane GT tworzone dla modułu PED	21
2.7	Przykładowe dane GT tworzone dla modułu MOD	23
2.8	Przykładowe dane GT tworzone dla modułu TSR i TLR	23
2.9	Przykładowe dane GT tworzone dla linii na drodze	25

2.10	Odcinki wyznaczające dane referencyjne dla granic wolnej przestrzeni (czerwony) oraz punkty będące wynikiem przetwarzania obrazu przez moduł percepcji (niebieski)	25
3.1	Prostokątne regiony rozpoznania oznaczające pieszych, samochody, znaki i sygnalizację świetlną	32
3.2	Przykłady obrazu otoczenia syntezowanego przez symulator jazdy CARLA [100], który jest uznawany za najlepszy pośród otwartych symulatorów przeznaczonych do rozwoju systemów autonomicznej jazdy [101]	40
4.1	Podobieństwo pozycji reprezentuje różnicę położenia prostokątów odniesioną do rozmiarów porównywanych obiektów	51
4.2	Wynik pomiaru jakości pozycji uzyskuje się po przeskalowaniu krzywej funkcji indywidualnie dla każdego porównania. Parametry p_1 i p_2 opisują główne progi oceny właściwości prostokątnych obszarów rozpoznania	52
4.3	Pierwszy sposób obliczania wag dla kolejnych indeksów sekwencji gdy $FD \leq CI$	58
4.4	Drugi sposób obliczania wag dla kolejnych indeksów sekwencji gdy $FD > CI$	59
4.5	Przykład sytuacji na drodze, kiedy spóźnione pierwsze wykrycie może odpowiadać pominięciu kluczowego zdarzenia podczas całego pojawienia się obiektu; 1 – włączenie się nowego obiektu do ruchu; 2 – podążanie za obiektem. Nowe podejście do oceny śledzenia zostało skonstruowane, tak aby uwrażliwić KPI na tę sytuację i odpowiednio podkreślić jej istotność podczas podsumowania ewaluacji	61
4.6	Porównanie analizy jakości śledzenia obiektu przy pomocy zaproponowanej metody, standardowej średniej oraz miary OSPA	63

4.7	Wizualizacja jakości rozpoznania kolejno pojawiających się w zasięgu czujników obiektów oceniona przy pomocy GMOS i miary podobieństwa sekwencji	65
4.8	Wizualizacja lokacji, klasyfikacji i opisu znalezionych obiektów fałszywie pozytywnych	67
4.9	Prostokąty referencyjne (czerwone) i prostokąty pochodzące z modułu percepcji PED (niebieskie - prawdziwie pozytywne i żółte - fałszywie pozytywne)	69
4.10	Prostokąty referencyjne (czerwone) i prostokąty pochodzące z modułu percepcji MOD (niebieskie)	71
4.11	Prostokąty referencyjne (czerwone) i prostokąty pochodzące z modułu percepcji TSR (niebieskie)	72
4.12	Prostokąty lidarowe (czerwone) i prostokąty radarowe (niebieskie - prawdziwie pozytywne i żółte - fałszywie pozytywne). Strzałki reprezentują wektor prędkości względnej obiektu	74
4.13	Wyniki eksperymentu Monte Carlo. Od lewej: porównanie rozkładów ewaluacji w klatce przy zaburzeniu pozycji, przy zaburzeniu rozrzutu oraz ewaluacja sekwencji przy zaburzeniu pozycji	79
5.1	Sześć grup scenariuszy testowych wydzielonych na podstawie warunków dominujących podczas sceny oraz ich symbole	85
5.2	Statystyki związane z procesem podejmowania decyzji o optymalnej liczbie podgrup na jakie zostanie podzielona grupa G3, podczas jej klastrowania. Optimum w tym przypadku zostało wybrane na podstawie metody łokciowej	88
5.3	Klastry scen z G2 wyznaczone przez algorytm. Na poziomej osi znajdują się numery scen, znaczniki reprezentują wartości miary podobieństwa sekwencji obiektów w tych scenach, a ich kolory wyznaczają klastry	89

5.4	Przykładowy fragment klatki filmu z wprowadzonymi różnymi poziomami rozmycia Gaussa. Od lewej - oryginał i rozmycia na poziomach 9, 27, 45 i 65	91
5.5	Przykładowy fragment klatki filmu po wyzerowaniu kolejno 0, 3, 8, 10, 11 ostatnich bitów rozdzielczości skali szarości	91
5.6	Podział klatki na segmenty, które zostały kolejno przesłonięte. W segmencie B1 znajduje się przykładowa elipsa imitująca fizyczne przesłonięcie sensora	92
5.7	Cykl oślepienia wprowadzony do filmu. Jeden cykl trwa 26 klatek, pierwsze $26 - k - 2s$ klatek jest niezmienionych, następnie w s klatkach poziom oślepienia narasta, przez k klatek jest na maksymalnym poziomie m , po czym przez s klatek spada do zera	92
5.8	Wizualizacja degradacji jakości GMOS (4.10) w klatkach przykładowego filmu. Na kolejnych segmentach widzimy tę samą scenę analizowaną przez moduł percepcji przy pogłębiającym się poziomie degradacji obrazu (rozmycie Gaussa)	94
5.9	Analiza wariancji w podgrupie G3S4 dla różnych rozkładów jakości rozpoznawania obiektów przy zwiększającym się poziomie rozmycia Gaussa	96
5.10	Analiza wariancji w podgrupie G1S1 dla różnych rozkładów jakości rozpoznawania obiektów przy zmniejszającej się rozdzielczości poziomów szarości	97
5.11	Analiza wariancji w podgrupie G5S3 dla różnych rozkładów jakości rozpoznawania obiektów przy okluzji obrazu obejmującej dziewięć różnych segmentów	98
5.12	Analiza wariancji w podgrupie G3S4 dla różnych rozkładów jakości rozpoznawania obiektów przy odmiennych typach cyklicznego oślepienia	99
5.13	Przykład wzrostu odległości Wassersteina dla postępującej intensywności rozmycia w grupie G3	107

5.14	Wartości Wassersteina dla postępującej intensywności rozmycia w grupie G1. Sztuczna degradacja nakłada się z naturalną	107
5.15	Przykład wartości Wassersteina dla przesłoniętych segmentów obrazu w grupie G6	108
5.16	Przykład wzrostu odległości Wassersteina dla różnych rodzajów cyklicznego oślepienia w grupie G5	108
5.17	Odległość Wassersteina między oryginalną jakością i jakością uzyskaną na filmach ze zredukowaną głębią bitową we wszystkich podgrupach eksperymentu	109
5.18	Przykład wzrostu odległości Wassersteina dla postępującej redukcji głębi bitowej w grupie G1	110
5.19	Wartości Wassersteina dla redukcji bitów w grupie G2. Sztuczna degradacja nakłada się z naturalną	110
6.1	Siatka zajętości przetransformowana z przestrzeni stanów \mathbb{X} zawartej w polu widzenia sensorów samochodu autonomicznego stanowiącego punkt odniesienia. Obiekty poruszające się względem niego podczas scenariusza testowego generują układy trajektorii	112
6.2	Przykładowe trajektorie zrekonstruowane dla danych radarowych, kropki oznaczają pojedyncze obserwacje pozycji dla obiektów poruszających się w otoczeniu sensora	113
6.3	Mapa termiczna zajętości otoczenia wokół samochodu znajdującego się w centrum układu współrzędnych i skierowanego w prawą stronę wykresu. Wykreślona ona została dla wszystkich trajektorii ze zbioru scenariuszy testowych	117

6.4	Przykłady map termicznych zajętości bazujących na scenariuszach z bazy danych nuScenes. Odległości Wassersteina między tymi scenami są zaprezentowane w tabeli 6.4	120
6.5	Kumulatywne mapy termiczne zajętości otoczenia dla trajektorii pojazdów i pieszych zawartych w zbiorach nuScenes zebranych w krajach USA i Singapur	122

Spis tablic

4.1	Podobieństwo prostokątów pochodzących z referencji i modułu percepcji PED na rysunku 4.9	69
4.2	Podobieństwo prostokątów pochodzących z referencji i modułu percepcji MOD na rysunku 4.10	72
4.3	Podobieństwo prostokątów pochodzących z referencji i modułu percepcji TSR na rysunku 4.12	73
4.4	Porównanie funkcjonalności miar podobieństwa prostokątnych regionów roz- poznania	77
5.1	Podsumowanie wyników ANOVA dla wprowadzonego rozmycia o różnym natężeniu i korespondujące wartości Wassersteina	101
5.2	Podsumowanie wyników ANOVA dla obrazów o zredukowanym poziomie szar- ości	102
5.3	Podsumowanie wyników ANOVA dla obrazów ze sztuczną okluzją	103
5.4	Podsumowanie wyników ANOVA dla filmów z cyklicznym oślepieniem	104
6.1	Wartości podobieństwa (%) w sensie normy L^p dla trajektorii przedstawi- onych na rysunku 6.2	114
6.2	Wartości podobieństwa (%) w sensie Jaccarda dla trajektorii przedstawi- onych na rysunku 6.2	115
6.3	Wartości podobieństwa (%) w sensie miary Hamminga dla trajektorii przed- stawionych na rysunku 6.2	116

6.4	Odległości Wassersteina pomiędzy mapami zajętości scenariuszy ze zbioru danych nuScenes	119
6.5	Współczynniki różnorodności wyznaczone dla fragmentów zbioru scenariuszy nuScenes pochodzących z różnych krajów	125

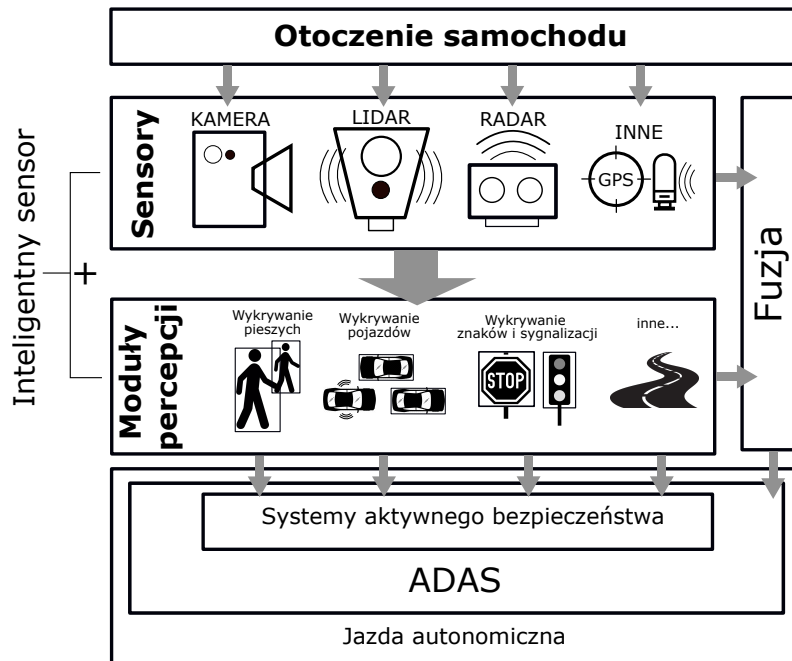
Rozdział 1

Wprowadzenie

Niniejsza praca stanowi próbę zmierzenia się z problematyką analizy danych związanych z rozwojem, testowaniem i weryfikacją podstaw działania inteligentnych systemów wspomaganie kierowcy w pojazdach. Celem pracy jest usprawnienie metodologii zautomatyzowanej analizy dużych ilości danych związanych z wybranymi aspektami ewaluacji modułów percepcji odpowiedzialnych za interpretowanie otoczenia w pobliżu samochodu. Moduły percepcji działają w oparciu o dane z różnych rodzajów sensorów i przetwarzają je dzięki zaawansowanym algorytmom wykorzystującym specjalizowane sztuczne sieci neuronowe. Na tej podstawie podejmowane są kluczowe decyzje związane z protokołami sterowania i bezpieczeństwa. Z tego powodu proces ewaluacji powinien uwzględniać złożoność przetwarzanych danych i ich przeznaczenie oraz powinien pozwalać na skuteczną lokalizację problemów. Proces rozwoju i testowania tego typu technologii w zróżnicowanych warunkach jest czasochłonny i kosztowny [1]. Dlatego przetwarzanie zebranych danych i analiza scenariuszy testowych powinna pozwalać na wyciąganie w sposób zautomatyzowany możliwie szczegółowych wniosków wyczerpujących potencjał owych danych.

1.1. Droga do autonomii

Bezpieczne kierowanie pojazdem wymaga nie tylko przestrzegania obowiązujących przepisów, ale również elastycznego dostosowania się kierującego do bieżącej sytuacji na drodze. Składają na nią z jednej strony nieruchome znaki drogowe (pionowe i poziome), sygnalizacja świetlna i droga, a z drugiej strony inni uczestnicy ruchu, w tym piesi. Owa elastyczność tyczy się zarówno kierującego (człowieka) jak i systemów wspierających go w prowadzeniu pojazdu. A zatem technologia pozwalająca na percepcję i interpretację



Rysunek 1.1. Diagram ilustrujący uproszczony schemat działania samochodu wyposażonego w inteligentną percepcję

w czasie rzeczywistym otoczenia samochodu jest istotnym składnikiem każdego systemu asystującego kierowcy.

1.1.1.1. Percepcja samochodu

Uproszczony schemat pokazujący istotne elementy i procesy stanowiące o istocie inteligentnego samochodu przedstawia rysunek 1.1. Na szczycie schematu znajduje się otoczenie, które jest obserwowane i interpretowane, aby algorytmy sterujące pojazdem mogły działać w oparciu o to co dzieje się na drodze.

Sensory [2]

Człowiek, aby zebrać informacje o otoczeniu w jakim się aktualnie znajduje, posługuje się pięcioma zmysłami. Samochód, aby mieć możliwość wychwytywania sygnałów z otoczenia, rozpoznawania ich i rejestracji również potrzebuje zbioru odpowiednich przyrządów pomiarowych – sensorów. Aby móc ciągle i kompleksowo zbierać dane o otoczeniu i zachodzących w nim zmianach na pokładzie pojazdu autonomicznego umieszcza się zwykle zestawy kamer rejestrujące widok z każdej strony pojazdu, radary krótkiego, średniego i dużego zasięgu, lidary, mikrofony i odbiorniki GPS. Tak jak nasze zmysły spisują się

w odbieraniu określonych bodźców, różne rodzaje sensorów są wyspecjalizowane oraz posiadają charakterystyczne mocne i słabe strony dlatego tylko ich dywersyfikacja pozwala na optymalizację wiarygodności ich wyników. Wymienione czujniki dostarczają ciągłego strumienia informacji o otoczeniu. Stanowi on absolutną podstawę do identyfikacji kontekstu aktualnej sytuacji drogowej w jakiej znajduje się pojazd, a to z kolei jest podstawą do wypracowania adekwatnej reakcji zaawansowanych systemów wspomaganie kierowcy samochodu. Jednak informacje płynące z sensorów to zaledwie początek. Aby systemy wspierające kierowcę mogły działać skutecznie nie wystarczy bowiem sama wiedza o tym, że otoczenie istnieje.

Moduły percepcji [3]

Surowe dane pobrane przez sensory muszą zostać odpowiednio zinterpretowane i przekształcone, tak aby być „zrozumiałymi” dla algorytmów sterowania. Analiza tego typu jest przeprowadzana w dużej mierze przez wiele wyspecjalizowanych sieci neuronowych stanowiących główny element modułów percepcji. Takie sieci odpowiadają za wyłowienie ze strumienia danych z sensorów wszystkich typów obiektów, które są istotne z punktu widzenia sprawnego prowadzenia bezpiecznej jazdy i dokładne ich opisanie. Każdy moduł percepcji w pojeździe jest wyspecjalizowany w konkretnym kierunku. Mają za zadanie identyfikować i szczegółowo scharakteryzować inne pojazdy na drodze, rowery, pieszych, pionowe i poziome znaki drogowe, sygnalizacje świetlne, linie, krawędzie drogi i inne istotne z punktu widzenia kierowania pojazdem okoliczności, jak na przykład obszar gdzie prowadzone są roboty drogowe. Ponieważ liczba możliwych sytuacji podczas ruchu drogowego jest w praktyce nieograniczona, a wymagany czas reakcji jest zwykle krótki, to podejmowane przez system bezpieczeństwa lub kierowania samochodem decyzje muszą bazować na wysoce adekwatnym, szczegółowym i wiarygodnym opisie otoczenia. Aby temu sprostać, moduły percepcji oparte na zaawansowanych sieciach neuronowych muszą być trenowane, testowane i weryfikowane na jak najbardziej zróżnicowanych i reprezentatywnych scenariuszach drogowych pozwalając im nauczyć się dobrze uogólniać problemy i wykazywać się adaptacją do nietypowych sytuacji. Oznacza to, że prawidłowo reagować powinny również w mniej standardowych sytuacjach. Sensor rozszerzony o dedykowany moduł percepcji nazywa się inteligentnym sensorem.

Fuzja [4]

Fuzja jest procesem, w którym wyniki algorytmów interpretujących dane pochodzące z różnych czujników i różnych modułów percepcji są łączone tak, aby uzyskać jeden spójny

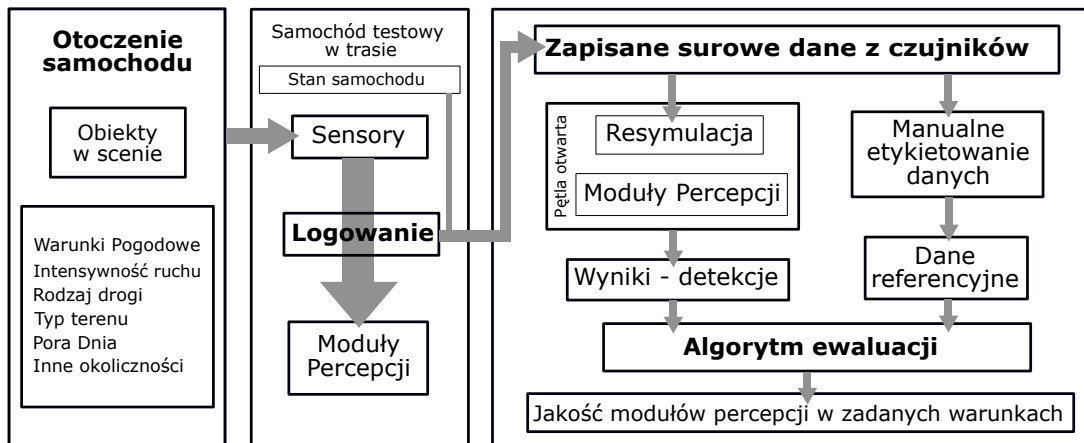
model otoczenia wokół pojazdu. Przykładowym zadaniem fuzji jest ustalenie czy przechodzień rozpoznawany przez algorytmy śledzące pracujące na danych radarowych jest tym samym pieszym oznaczonym przez sieć neuronową interpretującą dane z kamery przedniej.

1.1.2. Aktywne systemy bezpieczeństwa

Wyniki działania modułów percepcji są przetwarzane przez komputer pokładowy. W oparciu o nie mogą zostać podjęte różne akcje w zależności od rodzaju funkcji jakie pojazd posiada w ramach zaawansowanego systemu wspomagania kierowcy ADAS [5] (ang. *Advanced Driver Assistance System*) determinującego podstawowe poziomy autonomii pojazdu (więcej w sekcji 2.1.2). Do typowych funkcji działających praktycznie bezpośrednio na dyskretnych flagach ustawianych przez moduły percepcji należą systemy aktywnego bezpieczeństwa [6]. Umożliwiają one między innymi sygnalizowanie kierowcy za pomocą diod na desce rozdzielczej lub komunikatów dźwiękowych potencjalnego zagrożenia lub nawet automatyczne aktywowanie hamulców pojazdu w sytuacji bezpośredniego zagrożenia. Bardziej zaawansowane funkcje pojazdów autonomicznych to adaptacyjny tempomat, wspomaganie utrzymania pojazdu w świetle drogi, samodzielne parkowanie i inne [7]. Rozwój systemów autonomicznej jazdy stawia duże wymagania sprzętowe dla komputerów pokładowych, które na bieżąco muszą przetwarzać dane z wielu inteligentnych sensorów i podejmować decyzje o interakcjach z otoczeniem i kierowcą biorąc pod uwagę ich potencjalne skutki. Zdefiniowanie polityki [8], stworzenie i optymalizacja algorytmów autonomicznego systemu kierowania pojazdem stanowi kolejną szeroką dziedzinę badań, w których duże znaczenie mają metody uczenia maszynowego [9, 10].

1.1.3. Gromadzenie danych [11]

W celu przeanalizowania poprawności działania modułów percepcji potrzebne są dane z sensorów. Najlepiej aby dane te pobrane były w takich warunkach, w jakich przyjdzie w przyszłości tym modułom pracować. A zatem dane te powinny pochodzić z sensorów zamontowanych na poruszającym się pojeździe biorącym udział w różnych scenariuszach mogących mieć miejsce w prawdziwym ruchu drogowym. Aby zebrać takie dane przygotowuje się flotę samochodów testowych wyposażonych w zestawy opisanych wcześniej sensorów. Różnica w stosunku do typowego samochodu polega na tym, że dane z sensorów nie tylko są interpretowane ale też zapisywane na dyski twarde w jakie wyposażone są dodatkowo pojazdy testowe. Taka jazda testowa nie odbywa się w sposób autonomiczny, ale jest prowadzona przez zawodowych testerów. Aby zmaksymalizować szanse na zebra-



Rysunek 1.2. Diagram ilustrujący proces testowania modułów percepcji od momentu zebrania danych testowych, ich etykietowania do ewaluacji wyników uzyskanych przez resymulację

nie reprezentacyjnego zbioru danych flota jest rozsyłana na wszystkie kontynenty, tak aby w jak najbardziej zróżnicowanych warunkach zapisać w logach pojazdu dane zebrane przez sensory. Jazdy planowane są tak, aby spełnić założoną dystrybucję scen mających miejsce w określonych, zróżnicowanych warunkach pogodowych, typach terenu, rodzaju drogi, porze dnia, przy różnych poziomach natężenia ruchu, w miastach, na autostradach i terenach o innych poziomach urbanizacji. Dane te podczas logowania w pojeździe są synchronizowane i opatrywane informacją o stanie pojazdu testowego, na przykład o jego prędkości, która również jest częścią wymaganej dystrybucji warunków. Dane zgromadzone podczas jazd testowych umożliwiają inżynierom badanie działania modułów percepcji w idealnie powtarzalnych warunkach laboratoryjnych.

1.1.4. Testowanie i analiza jakości percepcji [1]

W procesie rozwoju modułów percepcji istotne jest sprawdzenie jak sobie one radzą z interpretacją danych z sensorów. Taka weryfikacja musi być przeprowadzona z dużą regularnością i zawsze gdy do ich kodu wprowadzane są istotne zmiany. Niepraktyczne jest tak częste i szczegółowe testowanie modułów percepcji z wykorzystaniem samochodów podróżujących po świecie. Dlatego właśnie dane zbierane przez sensory są podczas jazd testowych zapisywane. Po powrocie do laboratorium można je wykorzystać w procesie resymulacji (ReSim), który polega na skoordynowanym odtworzeniu uporządkowanego w czasie strumienia informacji z czujników i wstrzyknięciu ich na wejście testowanej właśnie wersji danego modułu percepcji. W taki sposób uzyskuje się w laboratorium wynik

ich działania, będący opisem poszczególnych obiektów, których obecność manifestuje się w danych zebranych przez sensory. Przykładowo, mogą to być przechodnie lub znaki nagrane na klatkach filmu pochodzącego z kamery lub chmury detekcji radarowych składające się na samochód jadący po sąsiednim pasie. Proces odtwarzania danych z czujników i uzyskiwania na ich podstawie wyników modułów percepcji w procesie ReSim nie jest objęty pętlą sprzężenia zwrotnego, ponieważ uzyskujemy jedynie informacje o detekcjach opisujących oryginalnie nagrane obiekty. Stanowią one zaledwie pierwszy z dwóch elementów potrzebnych do przeprowadzenia ewaluacji modułu percepcji, co pokazano na rysunku 1.2. Aby uczenie maszynowe i weryfikacja jego efektów mogła przebiegać prawidłowo należy uprzednio dokładnie opisać co w zebranych przez sensory danych powinno zostać znalezione i zinterpretowane przez moduły percepcji. Konieczne jest stworzenie danych referencyjnych – etykiet załączonych do danych z sensorów – opisujących oczekiwane wyniki z modułów percepcji. Zajmuje się tym sztab odpowiednio przeszkolonych osób, które manualnie analizują zebrane dane i etykietują je. W oparciu o te dodatkowe etykiety możliwe jest definitywne określenie jakości wyników otrzymanych z testowanego modułu percepcji, co dalej umożliwi doskonalenie go i ocenę skuteczności działania w warunkach rzeczywistych.

1.2. Problem badawczy

1.2.1. Motywacja

Proces gromadzenia danych przez samochody testowe mające na pokładzie zestawy nowoczesnych sensorów, które w każdym momencie rejestrują ogromne ilości danych sprawił, że wielkim wyzwaniem stało się ich przetwarzanie i składowanie danych [1]. O szybkości akumulowania danych dotyczących otoczenia samochodu testowego świadczy to, że przepustowość magistral w takim samochodzie obsługujących kilkanaście sensorów jest w pełni wykorzystywana, a pojedyncze dyski twarde nie są w stanie nadażyć z zapisem informacji, aby w sposób bezstratny je utrwalić. Ilość przestrzeni dyskowej potrzebnej do obsługi danych związanych z typowym projektem wykorzystującym AI w samochodach ciągle rośnie w tempie wykładniczym i dziś już przekracza 30 petabajtów ($30 \cdot 10^{15}$ bajtów) [1]. Aby takie ilości danych testowych mogły być użyteczne z punktu widzenia weryfikacji modułów percepcji muszą być poddane manualnemu etykietowaniu zgodnie ze szczegółowo skodyfikowanymi zasadami, co jest pracą mozolną i czasochłonną [1]. Ołbrzymia ilość pracy i zasobów poświęconych na zdobycie, przechowywanie i przygotowanie

danych sprawia, że koniecznością staje się opracowanie metody, która pozwoli je jak najlepiej spożytkować. Aby dokonać ewaluacji modułu percepcji potrzebna jest metodologia oceny jakości interpretacji otoczenia i poziomu odwzorowania danych referencyjnych przez moduł, ocena, która będzie możliwie adekwatna do przewidywanych zastosowań modułu. W celu obniżenia złożoności czasowej i sprzętowej takiego przedsięwzięcia oraz zagwarantowania powtarzalności wyników wymagana jest automatyzacja procesu analizy jakości. Jest to spowodowane tym, że w procesie rozwoju oprogramowania do interpretacji danych z sensorów testowanie jest często powtarzane nawet w cyklu trwającym zaledwie dwa tygodnie. Ponadto prawo wymaga aby zapewnić możliwość powtórzenia takiej analizy jeszcze po latach od zakończenia projektu, tak aby być w stanie potwierdzić deklarowaną jakość [1]. Badania nad rozwojem modułów percepcji inteligentnego pojazdu związane są z bezpieczeństwem ludzi i mają na celu zminimalizowanie liczby wypadków drogowych i szkód przez nie spowodowanych. Rozsądnym pomysłem wydaje się zatem stworzenie metodologii, która w rzetelny i obiektywny sposób oceni jakość prototypów oraz umożliwi szybką i skuteczną lokalizację problemów [11]. Aby wszyscy producenci pojazdów autonomicznych oraz prawodawcy mogli posługiwać się w tym zakresie uniwersalnym zbiorem zaleceń i dobrych praktyk powstają wymogi takie jak np. SOTIF [12] (ang. *Safety Of The Intended Functionality*). Postuluje się w nich konieczność potwierdzenia skuteczności systemów ADAS. Liczba sytuacji w jakich może znaleźć się samochód autonomiczny jest w praktyce nieograniczona. Aby to uwzględnić scenariusze testowe powstają w mocno zróżnicowanych warunkach i zawierają złożony opis różnego rodzaju obiektów biorących aktywny udział w przebiegu ruchu na drodze. Dlatego istotnym jest zaprojektowanie podejścia, które pozwoli na syntezę szczegółowych wniosków wyczerpujących potencjał jak największej ilości zebranych danych. Na etapie projektowania eksperymentów możliwość oceny jakie dane są szczególnie użyteczne i w jakiej ilości muszą być zebrane by stanowić wystarczająco dobrą reprezentację rzeczywistości w sposób naturalny przekłada się na oszczędność zaangażowanych środków [1]. Jest to ściśle związane z procesem ewaluacji wyników dlatego metoda oceny nie tylko jakości percepcji ale również jakości i różnorodności danych testowych jest bardzo cenna z punktu widzenia całego przedsięwzięcia [13].

1.2.2. Tezy rozprawy

1. Możliwym jest stworzenie metodologii ewaluacji, która precyzyjniej określa jakość wyznaczenia prostokątnych regionów rozpoznania niż robi to indeks Jaccarda w kontekście detektorów wizyjnych wchodzących w skład modułów percepcji. Ponadto

można stworzyć miarę podsumowującą całe pojawienie się obiektu w zasięgu sensorów uwzględniającą potencjalną konieczność szybkiej reakcji na nową sytuację.

2. Usystematyzowanie procesu testowania modułów percepcji w zróżnicowanych warunkach ruchu drogowego i istotne rozszerzenie płynących z niego wniosków można uzyskać poprzez uzupełnienie metody testowania hipotez statystycznych o analizę topologiczną przeprowadzoną z wykorzystaniem algorytmu klastrowania i odległości Wassersteina. Dodatkowo odległość Wassersteina może rozszerzyć samą metodologię oceny jakościowej.
3. Analiza zajętości obszarów znajdujących się w zasięgu sensorów samochodu testowego z wykorzystaniem metryki Wassersteina do klastrowania dwuwymiarowych rozkładów jest metodologią umożliwiającą formalizację opisu danych uczących, testowych i walidacyjnych używanych do szeroko rozumianego rozwoju jazdy autonomicznej oraz pozwoli ocenić ich różnorodność.

1.2.3. Granice zakresu rozważań

Niniejsza praca doktorska dotyczy zagadnień związanych z testowaniem i walidacją wyników oprogramowania związanego z inteligentną percepcją w pojazdach, a dokładniej z analizą gromadzonych i etykietowanych danych testowych służących do opisu scenariusza testowego oraz do weryfikacji i walidacji modułów percepcji. Badane w pracy algorytmy ewaluacji i metodologie analizy struktury danych testowych przysłużyć się mają inżynierom pracującym nad rozwojem systemów aktywnego bezpieczeństwa. Spośród modułów percepcji skupiono się na tych, które analizują obiekty nadające się do opisu przy pomocy prostokątnych regionów rozpoznania. Większość prac dotyczy oceny interpretacji obrazu z kamery przedniej stanowiącej źródło najważniejszych informacji o środowisku w jakim pojazd się znajduje aktualnie i w jakim znajdzie się podczas jazdy w najbliższej przyszłości. W dalszej części pracy przetwarza się również interpretację scenariusza testowego, którą można uzyskać przy pomocy modułów pracujących na podstawie pozostałych sensorów. Tematyka pracy nie dotyczy procesu projektowania sprzętu, w szczególności sensorów. Również fuzja pomimo roli podobnej do modułów percepcji nie jest rozważana w tej pracy. Praca nie skupia się na porównywaniu wyników różnych sieci neuronowych lecz na porównaniu różnych metod analizy jakości ich działania oraz sposobów przetwarzania danych związanych z ich testowaniem w kontekście wykorzystania w zaawansowanych systemach wspierania kierowcy w pojeździe.

1.2.4. Zagadnienia omawiane w rozdziałach

Rozdział drugi omawia z czego jest zbudowany, jak działa i jak jest tworzony system percepcji w pojazdach autonomicznych ze szczególnym uwzględnieniem sensorów znajdujących się na pokładzie tego typu pojazdu oraz modułów percepcji stanowiących podstawę dla interpretacji przez komputer otaczającego go świata. Rozdział trzeci przybliży proces ewaluacji działania modułów percepcji w samochodzie wyposażonym w systemy aktywnego bezpieczeństwa oraz stanowi przegląd literatury dotyczący metod analizy danych charakterystycznych dla testowania inteligentnej percepcji dla systemów asystujących kierowcy. Dotyka też możliwości jakie dają różne techniki wizualizacji i analizy danych za pomocą metryki Wassersteina. W rozdziale czwartym zaprezentowano autorskie miary ewaluacji modułów percepcji pracujących na bazie prostokątnych regionów rozpoznania. Metodologia oceny zawiera analizę lokalnego podobieństwa w klatce jak i podsumowanie ewaluacji wykrycia obiektów w czasie. Zaprezentowane metody są również omawiane na konkretnych przykładach kalibracji dla wybranych modułów percepcji. Opisana metodologia została opublikowana w pracy [14]. Rozdział piąty opisuje eksperyment mający na celu zbadanie wpływu degradacji danych wizyjnych na jakość działania modułu percepcji po wprowadzeniu sztucznych zaburzeń do obrazu z kamery. Na podstawie tego obrazu badany moduł rozpoznaje poruszające się na drodze samochody. Wykorzystano tutaj opisane w poprzednim rozdziale narzędzia oraz zaproponowano metodologię kreowania układu eksperymentalnego wykorzystującą klastrowanie porównywanych rozkładów jakości z wykorzystaniem metryki Wassersteina. Wyniki te udało się opublikować w [15] oraz [16]. Pokazano też w jaki sposób obróbka statystyczna wyników eksperymentu może zostać rozszerzona o topologiczną analizę porównawczą jakości wykorzystującą koncepcję odległości Wassersteina. W rozdziale szóstym omówiono sposób analizy dużych zbiorów danych zawierających scenariusze testowe składające się ze zbiorów trajektorii opisanych w układzie współrzędnych, w którym centrum stanowi samochód testowy. Zaproponowano metodę charakteryzacji i porównywania takich zbiorów z wykorzystaniem termicznej mapy zajętości otoczenia oraz odległości Wassersteina dwuwymiarowych rozkładów zajętości generowanych przez zawarte w scenariuszach trajektorie obiektów. Wyniki znajdują się również w publikacjach [17] i [18]. Ostatni rozdział zawiera podsumowanie przeprowadzonych prac i jego wkładu w rozwój metod ewaluacji modułów percepcji tworzonych dla samochodów wykorzystujących systemy ADAS.

Rozdział 2

Budowa i testowanie inteligentnej percepcji pojazdu

2.1. Pojazd autonomiczny

Obecnie jedną z najbardziej dynamicznie rozwijających się gałęzi motoryzacji są technologie związane z jazdą autonomiczną. W słowniku języka polskiego autonomia to samodzielność i niezależność w decydowaniu o sobie. Docelowo jazda autonomiczna ma doprowadzić nie tylko do zmniejszenia liczby kierowców na drogach ale przede wszystkim do zmniejszenia poziomu zagrożenia związanego z ruchem na drodze oraz wielu innych negatywnych czynników wynikających z nieustannie rosnącego zapotrzebowania na transport ludzi i towarów.

2.1.1. Definicja

Samochodem autonomicznym jest pojazd zdolny do zidentyfikowania swojego otoczenia i bezpiecznego poruszania się w nim bez ingerencji kierowcy. Systemy autonomiczne charakteryzują się możliwością podejmowania decyzji niezależnie od człowieka [19, 20], ale proces ten jest rozwinięciem zwykłej automatyzacji przez fakt, że decyzyjność jest możliwa pomimo złożonej niepewności związanej z innymi uczestnikami ruchu [21].

Motywacji do rozwoju ADAS oraz autonomicznej jazdy jest wiele. Najważniejszym jest aspekt związany z bezpieczeństwem. Dzisiaj ogromna większość bo ponad 90% poważnych wypadków na drodze spowodowana jest błędem kierowcy [22]. Wysoki poziom inteligentnej percepcji i systemów wspomagania w pojazdach zredukuje ryzyko i niebezpieczeństwo na drodze wynikające z rozproszenia uwagi lub zmęczenia kierowcy, nadmiernej prędkości,

słabej widoczności lub jazdy pod wpływem środków odurzających. Samochody takie mogą ułatwić bezpieczne i skuteczne korzystanie z pojazdu osobom z niepełnosprawnością oraz osobom starszym. Przekłada się to dodatkowo na oszczędność pieniędzy potrzebnych na leczenie ofiar wypadków (nie wspominając o ich cierpieniach fizycznych i emocjonalnych), naprawy uszkodzonych pojazdów i struktur drogowych. Wysokie poziomy autonomiczności pojazdów otwierają dla kierowcy możliwość bardziej produktywnego wykorzystania czasu podróży. Nie dotyczy się to tylko podróży w samosterującym się pojeździe, podczas której można podjąć inne zajęcia niż kierowanie pojazdem. Bardziej pożytecznie wykorzystany zostanie czas gdy będzie można wysiąść z samochodu u celu i pozostawić mu zadanie poszukiwania miejsca na parking. Powszechność rozwiązań autonomicznych w samochodach przyczyni się do redukcji korków, zwiększenia płynności ruchu drogowego [23, 24] i bardziej ekonomicznego wykorzystania paliwa co z kolei przekłada się również na zmniejszenie szkodliwych emisji do atmosfery.

Stan prawny w Polsce i UE

Według ustawy z dnia 20 czerwca 1997 roku dotyczącej *prawa o ruchu drogowym* pojazdem autonomicznym jest pojazd samochodowy, wyposażony w systemy sprawujące kontrolę nad ruchem tego pojazdu i umożliwiające jego ruch bez ingerencji kierującego [25].

Rozporządzenie Parlamentu Europejskiego i Rady Unii Europejskiej z dnia 27 listopada 2019 roku [26] *w sprawie wymogów dotyczących homologacji typu pojazdów silnikowych i ich przyczep oraz układów, komponentów i oddzielnych zespołów technicznych przeznaczonych do tych pojazdów, w odniesieniu do ich ogólnego bezpieczeństwa oraz ochrony osób znajdujących się w pojeździe i niechronionych uczestników ruchu drogowego* wprowadza nowe obowiązkowe funkcjonalności bezpieczeństwa jakie mają być zawarte w pojazdach od roku 2022. Ma to być dla samochodów osobowych, dostawczych, ciężarowych i autobusów: ostrzeżenie o senności kierowcy i rozproszeniu uwagi (np. korzystanie ze smartfona podczas jazdy), inteligentne wspomaganie prędkości, cofanie wspomagane kamerą lub innymi czujnikami oraz rejestrator danych w razie wypadku. Dla samochodów osobowych i dostawczych: asysta w utrzymaniu pasa ruchu, zaawansowane hamowanie awaryjne i ulepszone pasy bezpieczeństwa w testach zderzeniowych. W niedalekiej przyszłości obowiązkowym mają się również stać rozwiązania mające na celu wsparcie percepcji pieszych i rowerzystów dla kierowców autobusów i ciężarówek oraz usuwanie martwych pól z przodu i z boku tych pojazdów.

Systemy aktywnego bezpieczeństwa

Systemy aktywnego bezpieczeństwa to termin odnoszący się do samochodowych technologii zaprojektowanych w celu zmniejszenia prawdopodobieństwa wypadku, kolizji lub innego niebezpiecznego incydentu na drodze. Aktywne środki bezpieczeństwa obejmują technologię, która ma zapewnić bezpieczniejsze zachowanie podczas jazdy, ostrzegać kierowców o potencjalnych zagrożeniach i automatycznie się aktywować w przypadku, gdy poziom zagrożenia staje się zbyt wysoki [27]. Systemy bezpieczeństwa aktywnego jak na przykład elektroniczna kontrola stabilności i automatyczne hamowanie awaryjne są zasadniczo związane z zapobieganiem lub minimalizacją dotkliwości wypadków przed ich nastąpieniem w przeciwieństwie do pasywnych systemów bezpieczeństwa takich jak poduszki powietrzne działających w reakcji na wypadek.

2.1.2. Klasy autonomiczności

Bazując na taksonomii [28] zaproponowanej przez SAE (*Society of Automotive Engineers*) stanowiącej bazę dla standardu ISO/SAE PAS 22736:2021 należy rozróżnić następujące poziomy autonomiczności

- **Poziom 0: Brak autonomiczności** – kierowca jest odpowiedzialny za wszystkie zadania związane ze sterowaniem pojazdem oraz obserwacją otoczenia.
- **Poziom 1: Wspomaganie kierowcy** – samochód może kontrolować prędkość albo sterowanie pojazdem w ograniczonych warunkach asystując kierowcy. Oczy kierowcy przez cały czas muszą obserwować drogę, a ręce znajdować się na kierownicy.
- **Poziom 2: Częściowa automatyzacja** – samochód może jednocześnie kontrolować prędkość i sterowanie pojazdem w ograniczonych warunkach asystując kierowcy. Oczy kierowcy przez cały czas muszą obserwować drogę, ręce kierowcy nie muszą być trzymane na kierownicy w określonych sytuacjach.
- **Poziom 3: Warunkowa automatyzacja** – samochód może poruszać się autonomicznie w standardowych warunkach ale wymaga nadzoru i gotowości do interwencji kierowcy. Oczy kierowcy nie muszą obserwować drogi, a ręce nie muszą znajdować się na kierownicy ale kierowca musi być gotowy w każdej chwili przejąć panowanie nad pojazdem jeśli zostanie to zasygnalizowane.
- **Poziom 4: Wysoka automatyzacja** – samochód może poruszać się autonomicznie bez nadzoru kierowcy ale w standardowych warunkach. Kierowca nie musi obser-

wować drogi, trzymać ręki na kierownicy ani być gotowym do przejęcia sterów bo samochód będzie w stanie w sytuacji wyjątkowej powrócić do bezpiecznego stanu sam.

- **Poziom 5: Pełna autonomia** – pojazd porusza się autonomicznie w każdych warunkach. Brak kierownicy w pojeździe.

Analizą gotowości do przejęcia kontroli nad pojazdem zajmują się kamery DMS (ang. *Driver Monitoring Systems*) [29]. Znajdują się one wewnątrz pojazdu i obserwują jego wnętrze. Sprawdzają one czy ręce kierowcy znajdują się na kierownicy, kiedy jest to wymagane i czy jego oczy obserwują drogę gdy jest to konieczne. Dodatkowo algorytmy pracujące na danych z kamery mogą oceniać czy kierowca jest senny lub rozproszony w inny sposób. W sytuacji gdy kierowca nie jest gotowy pełnić odpowiednich obowiązków samochód może informować o tym przez sygnał dźwiękowy lub przez sygnalizator na desce rozdzielczej. Nieformalnie poziomy 1 i 2 uznaje się za domenę systemów ADAS, natomiast dopiero poziomy 3 i wyżej nazywa się jazdą autonomiczną (AD) (ang. *Autonomous Driving*) [30]. Rozwój legislacji w tym obszarze jest bardzo potrzebny i wciąż trwa na całym świecie. Dla skrócenia zapisu w poniższej pracy za samochód autonomiczny uznaje się samochód o niezerowym poziomie autonomii, posiadający przynajmniej systemy aktywnego bezpieczeństwa wymagające modułów percepcji. Potrzebne są one również na wyższych poziomach i na ich testowaniu praca się skupia.

2.1.3. Pojazdy autonomiczne dostępne na rynku

Aktualnie na rynku istnieje duża reprezentacja samochodów z wyposażeniem pozwalającym na sklasyfikowanie ich jako posiadające poziom 2 autonomiczności. Przykłady pojazdów produkowanych seryjnie (stan na rok 2022) można zobaczyć na rysunku 2.1 i od lewej są to:

- Tesla Model S (poziom 2) [31],
- Audi A6 Limousine (poziom 2) [32],
- Audi RS Q8 (poziom 3) [33].

Pojazdy autonomiczne to właściwie roboty. Podobne technologie są również wykorzystywane między innymi w dronach, czołgach, maszynach rolniczych, wózkach widłowych itp. Sama technologia percepcji i analizy otoczenia na bazie sensorów, które są wykorzystane



Rysunek 2.1. Przykłady produkowanych seryjnie modeli samochodów o niezerowym poziomie autonomiczności

w samochodzie autonomicznym ma ogromne zastosowanie przykładowo w automatycznym sterowaniu ruchem na drodze oraz monitoringu.

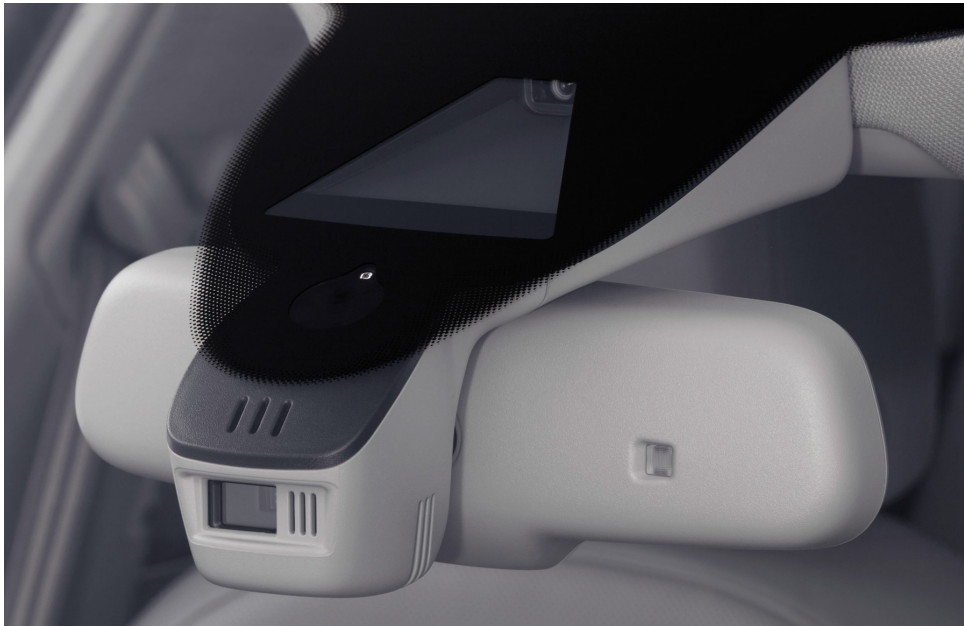
2.2. Samochodowe „zmysły”

Czujniki takie jak kamery, radary i lidary są obecnie niezbędnymi elementami wielu samochodowych systemów sterowania. Surowe dane dostarczane przez te sensory stanowią dane wejściowe do modułów percepcji, których celem jest monitorowanie i interpretowanie tego, co dzieje się w bliskim i odległym otoczeniu pojazdu. Wyniki działania algorytmów analizujących dane z czujników są podstawą do automatycznego zrozumienia sytuacji na drodze co z kolei jest wykorzystane do dobrania optymalnej reakcji pojazdu.

2.2.1. Kamery

W kwestii zrozumienia środowiska najbliższym ludzkiemu sposobowi postrzegania rzeczywistości sensorem jest kamera. Samochody autonomiczne wyposażone są zwykle w zestaw różnego rodzaju kamer umieszczonych z każdej strony pojazdu rejestrujących na klatkach wideo otoczenie w jakim on się porusza. Niektóre mają szerokie pole widzenia — aż 120 stopni — i mniejszy zasięg. Inne, jak na przykład frontowa kamera (rysunek 2.2) skupiają się na węższym widoku, aby zapewnić wizualizację dalekiego zasięgu. Możliwym jest też wykorzystanie kamery typu „rybie oko”, która zapewnia panoramiczny widok, aby dać pełny obraz tego, co znajduje się za pojazdem, tak by mógł się sam zaparkować. Jak wspomniano wcześniej kamery są też umieszczone wewnątrz pojazdu w celu analizy stanu kierowcy. Kamery służące percepcji ruchu na drodze posiadają szereg istotnych parametrów wpływających na ich jakość działania między innymi:

- klatkaż – FPS (ang. *frames per second*) - duża ilość klatek rejestrowanych na se-

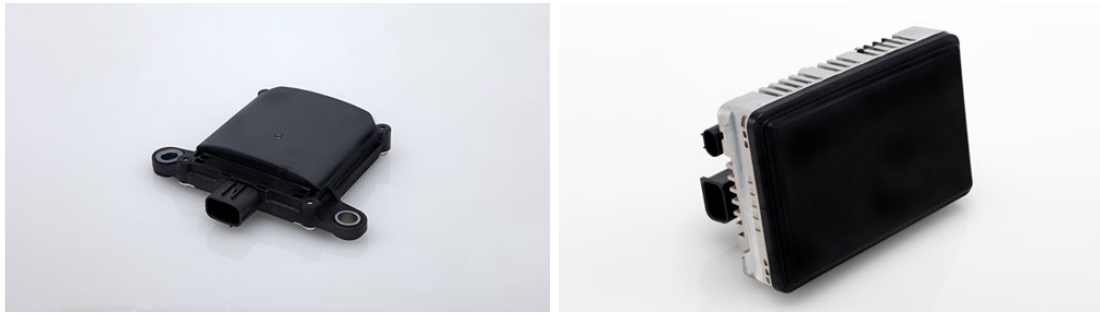


Rysunek 2.2. Kamera przednia wkomponowana w lusterko wsteczne samochodu marki Audi [34]

kundę zmniejsza szansę na szybką detekcję i nie pominięcie istotnych szczegółów szczególnie przy dużej prędkości pojazdu rozmywającej obraz.

- Rozdzielczość - wyższa rozdzielczość umożliwia bardziej precyzyjną klasyfikację obiektów znajdujących się w dużej odległości.
- Stosunek sygnału do szumu (SNR, ang. *signal-to-noise ratio*) oraz jasność mają wpływ na możliwość interpretowania obrazu przy niedostatecznym lub gwałtownie zmieniającym się oświetleniu (na przykład z boku pojazdu w nocy).

Sensory tego rodzaju są szczególnie dobre w zadaniach wymagających wysokiej rozdzielczości, takich jak klasyfikacja znaków drogowych lub zadania wymagające postrzegania kolorów jak w przypadku rozpoznania stanu sygnalizacji świetlnej. Są też stosunkowo tanie w porównaniu z radarem i lidarem. Do wad tego sensora można zaliczyć trudności w pozyskaniu precyzyjnych danych o odległości obiektów od samochodu. Aczkolwiek istnieją doniesienia o algorytmach, które są w stanie oszacować relatywnie dokładnie odległości obiektów od samochodu na podstawie zapisu wideo, na przykład te opisane w pracach [35] lub [36], to jednak w tym przypadku radar pozostaje standardem [37]. Tym bardziej, że rozpoznanie obiektów na obrazie z kamery skutecznie utrudniają złe warunki pogodowe takie jak np. deszcz i śnieg. Naturalnym utrudnieniem jest również brak lub nierównomierne oświetlenie na przykład w nocy lub podczas wjazdu i wyjazdu z tunelu [2].



Rysunek 2.3. Przykładowe modele radarów marki Continental przeznaczone dla samochodów autonomicznych. Radar krótkiego zasięgu po lewej [38] i radar długiego zasięgu po prawej [39]

Obrazy z kamer podobnie jak pozostałe sensory dzięki dedykowanym systemom na czipie SoC (ang. *system on a chip*) są analizowane przez różne algorytmy wspierane przez zaawansowane sieci neuronowe by zlokalizować wszystkie istotne obiekty znajdujące się w otoczeniu i mające wpływ na ruch drogowy.

2.2.2. Radary

Standardowe radary transmitują fale elektromagnetyczne i analizują czas powrotu do źródła fal odbitych aby znaleźć odległość od sensora obiektu odbijającego. Zwykły radar zamontowany na szybko przemieszczającym się pojeździe w otoczeniu, w którym wiele jego elementów jest w stanie odbić falę może mieć trudność w rozróżnieniu które informacje dotyczą ważnych w kontekście prowadzenia bezpiecznej jazdy obiektów odbijających. Szczególnie w scenie gdzie porusza się bardzo wiele obiektów w różnych kierunkach i o różnych prędkościach przykładowo na tle budynków. Dlatego w zadaniach percepcji samochodu szczególnie użyteczne są radary dopplerowskie. Tego typu radary są wysoce wrażliwe na zmianę dynamiki otoczenia. Oznacza to, że jeśli samochód się nie porusza, a otoczenie jest statyczne taki radar nie raportuje nic. Są one szczególnie użyteczne ze względu na możliwość odfiltrowania nieistotnego tła. Wykorzystuje się je nie tylko do estymacji pozycji poruszających się w ich zasięgu obiektów, umożliwiają one również wyznaczanie kierunku i wartości wektora prędkości względnej obiektu odbijającego [40]. Szczególnie istotnym wskaźnikiem wyliczanym w czasie rzeczywistym na podstawie danych radarowych jest TTC (ang. *Time to collision*) czyli czas do hipotetycznej kolizji obiektów w układzie o ustalonych parametrach dynamicznych [41]. Zbyt niski poziom TTC wyznaczanego względem obiektów w systemach aktywnego bezpieczeństwa umożliwia wysyłanie automatycznego ostrzeżenia kierowcy lub wprost aktywowanie hamulców

pojazdu. Stosowane obecnie systemy radarowe wyposażone są w algorytmy, które oprócz klastrowania, śledzenia oraz dostarczania informacji o odległości i prędkości celu, potrafią również klasyfikować niektóre obiekty. Radary stosowane w samochodach autonomicznych można podzielić ze względu na ich zasięg oraz funkcjonalność do jakiej są przypisane (rysunek 2.3).

- *Radary krótkiego zasięgu* – wykorzystywane na przykład przez asystenta parkowania.
- *Radary średniego zasięgu* – są wykorzystywane między innymi do detekcji ruchu odbywającego się w kierunku poprzecznym do ruchu samochodu lub alarmowania o szybko zbliżającym się od tyłu pojeździe.
- *Radary długiego zasięgu* – służą przede wszystkim do śledzenia obiektów w dużej odległości przed samochodem i są wykorzystane między innymi w adaptacyjnym tempomacie.

2.2.3. Lidary



Rysunek 2.4. Lidar zamontowany na dachu samochodu testowego na torze Aptiv [42]

Lidar, a właściwie LiDAR (ang. *Light Detection and Ranging*) działa na tej samej zasadzie co radar, ale zamiast mikrofal wykorzystuje impulsy świetlne (rysunek 2.4). Dzięki



Rysunek 2.5. Czujnik ultradźwiękowy [44] (po lewej) i system mikrofonów [45] (po prawej)

wysokiej rozdzielczości lidar może być używany do klasyfikowania obiektów oraz do mierzenia ich odległości od samochodu. Technologia lidarowa ze względu na koszty i trwałość nie jest jeszcze często wykorzystana w masowo produkowanych pojazdach. Aktualnie w większości przypadków linary są używane aby umożliwić precyzyjne etykietowanie i pozycjonowanie obiektów wokół pojazdu rejestrującego dane. Najbardziej popularne ze względu na jakość w stosunku do ceny są urządzenia obrotowe jednak w produkcji końcowym może to wyglądać inaczej [43] między innymi ze względu na konieczność optymalizacji gabarytów i prezencji tego czujnika. w skutecznym wykorzystaniu ich do analizy otoczenia samochodu decydującą rolę odgrywają parametry takie jak duży zasięg, wysoka rozdzielczość przestrzenna i wydajność w czasie rzeczywistym. Muszą też charakteryzować się tolerancją na słoneczne tło aby zapewnić skuteczność. Systemy lidarowe zapewniają znacznie dokładniejsze pomiary niż radarowe dlatego ich wyniki często są wykorzystane jako referencja przy testowaniu systemów radarowych. Jednakowoż zła pogoda jak deszcz, mgła czy śnieżyca znacznie bardziej obniża skuteczność lidarów niż radaru [2].

2.2.4. Czujniki ultradźwiękowe i mikrofony

Czujniki ultradźwiękowe znajdują swoje zastosowanie przy obliczaniu odległości do przeszkód podczas parkowania i manewrowania. Mogą wspierać funkcję awaryjnego hamowania przy niskich prędkościach ze względu na szybszy czas reakcji na niektóre przeszkody. Układy zbudowane z zestawu mikrofonów również mogą być zastosowane do

lokalizacji źródła dźwięku [46], [47] poprzez analizę różnicy dotarcia sygnału do różnych sensorów (rysunek 2.5).

2.3. Moduły percepcji

Aby planować tor ruchu i trzymać się drogi, dostosowywać sterowanie do znaków i sygnalizacji, odpowiadać szybko i skutecznie na wykonywane przez innych uczestników ruchu działania samochód autonomiczny musi posiadać możliwość percepcji otoczenia i interpretować na bieżąco środowisko w jakim się znajduje.

2.3.1. Definicja

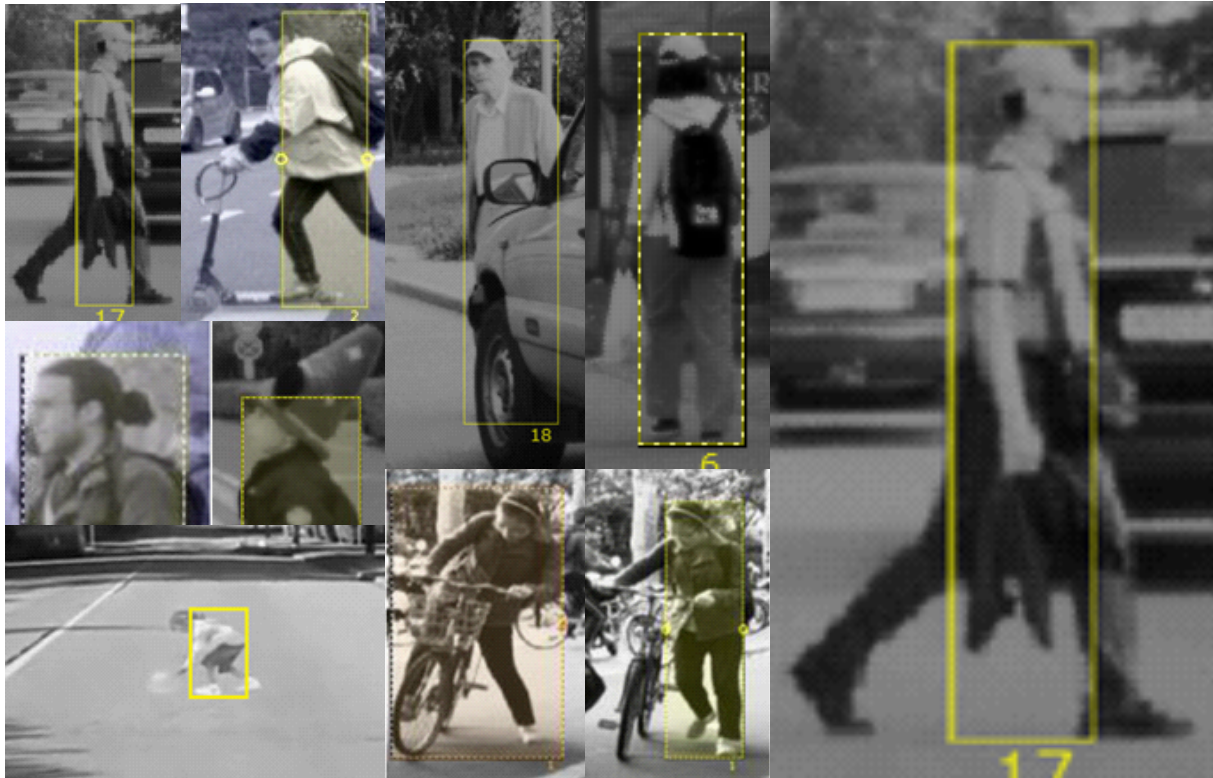
Moduł percepcji tworzą pakiety oprogramowania, które bazując na danych przechwyconych przez sensory interpretują różne aspekty otoczenia pojazdu. Moduł wykorzystuje moc obliczeniową komputerów pokładowych pojazdu, aby przetworzyć dane pochodzące od sensorów z wykorzystaniem odpowiednich algorytmów detekcji i śledzenia [48]. Wspomniane algorytmy oparte są przede wszystkim o wykorzystanie sztucznych sieci neuronowych do identyfikacji poszczególnych elementów drogi oraz jej otoczenia, które są istotne z punktu widzenia prowadzenia bezpiecznej jazdy. Moduł następnie publikuje listę obiektów w przestrzeni obejmowanej przez sensory w układzie współrzędnych, którego początek powiązany jest z pojazdem. Ponadto moduł odpowiada za określenie wybranych parametrów statycznych i dynamicznych wykrytych obiektów. Takie dane następnie w zależności od rodzaju systemów wspomagania kierowcy i poziomu autonomii pojazdu są przekazywane do odpowiednich algorytmów sterowania pojazdem. Proces ten może się różnić w zależności od szczegółów architektury oprogramowania samochodu [49], [50] ale idea pozostaje zawsze ta sama – na podstawie danych z sensorów moduł percepcji interpretuje otoczenie pojazdu.

2.3.2. Dane referencyjne

Dane referencyjne (ang. *Ground Truth* – GT) to termin używany w różnych dziedzinach w odniesieniu do informacji, o których zakłada się, że są zgodne z rzeczywistością [51]. Zwykle decyduje o tym fakt, że zostały uwierzytelnione przez eksperta, więcej o tym można przeczytać w sekcji 2.4.3. Z perspektywy testowania modułów percepcji nazywa się je również danymi wzorcowymi. Aby zweryfikować dany moduł percepcji należy opracować wymagania dotyczące przygotowania danych referencyjnych, na podstawie

których zostaną obliczone współczynniki jakości. W zależności od wymagań klienta takie reguły dotyczące opisywania rozpoznawanych przez moduł obiektów mogą się różnić w niektórych aspektach.

2.3.3. Moduł wykrywający pieszych



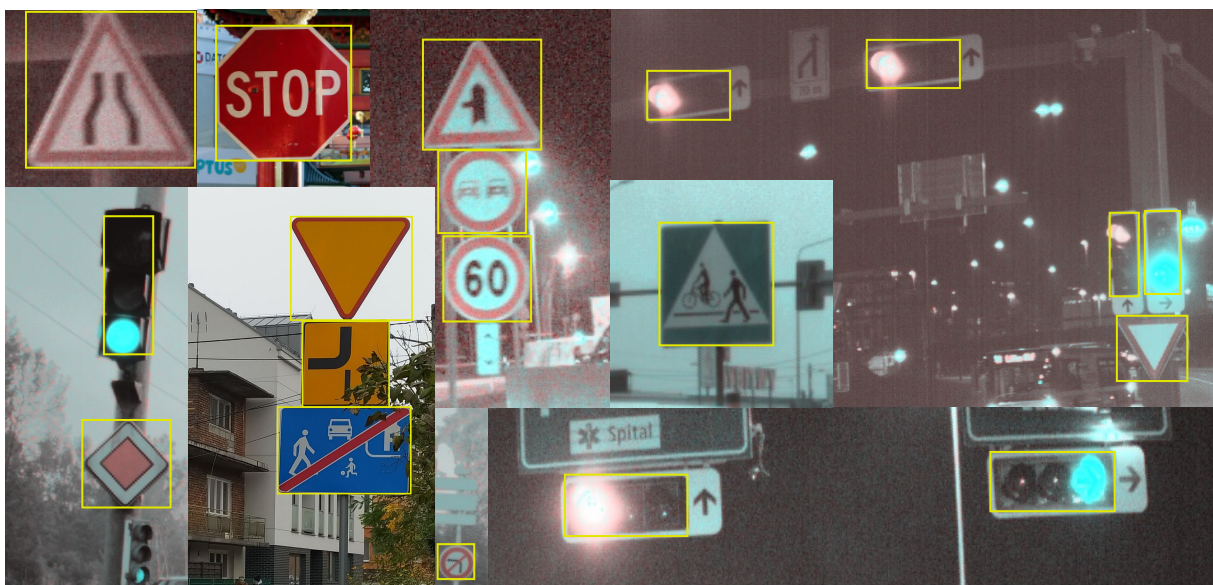
Rysunek 2.6. Przykładowe dane GT tworzone dla modułu PED

Moduł percepcji wykrywający pieszych - PED (ang. *pedestrian*) zajmuje się oznaczaniem przechodniów na klatkach filmów pochodzących z kamer zamontowanych w samochodzie. Konieczność wykrywania takich obiektów dotyczy nie tylko branży motoryzacyjnej, ale jest również częścią inteligentnego monitoringu związanego z bezpieczeństwem publicznym [52]. Przykładowa seria reguł opisujących jak przygotować dane wzorcowe dla modułu PED wyglądać może następująco. Oznaczamy standardowo pieszych, którzy stoją lub są pochyleni. Leżące postacie również są oznaczone ale z dodatkową flagą (informacją o stanie obiektu oznaczonego danym prostokątem). Postacie są etykietowane pod warunkiem, że prostokąt ograniczający będzie zawierał nie mniej niż ustaloną liczbę pikseli wysokości i szerokości. Dodatkowo obiekt powinien być widoczny (niezasłonięty) w przynajmniej 40%. W przypadku gdy zakrycie jest znaczące ale zawiera się w okre-

ślonych granicach tolerancji oznaczenie powinno obejmować również niewidoczną część obiektu. W zależności od wymagań można spotkać regułę zobowiązującą do oznaczania pieszych razem z nakryciem głowy lub bez. Odpowiednio oznaczać należy również rysunki ludzi pojawiających się na samochodach oraz perspektywiczne rysunki na jezdni. Ten moduł obejmuje również klasyfikowanie rowerzystów. Rowerzyści są oznaczani w całości razem z rowerem, osoba prowadząca rower jest pieszym, staje się rowerzystą w momencie przełożenia jednej nogi przez ramę roweru. Powyższe zasady to jedynie przykłady reguł tworzenia danych referencyjnych (rysunek 2.6). Ich złożoność sprawia, że manualne lub nawet półautomatyczne tworzenie danych GT jest czasochłonne i kosztowne dlatego istotnym jest by skutecznie je wykorzystywać.



Rysunek 2.7. Przykładowe dane GT tworzone dla modułu MOD



Rysunek 2.8. Przykładowe dane GT tworzone dla modułu TSR i TLR

2.3.4. Pojazdy na drodze

Moduł percepcji MOD (ang. *Mobile Object Detection*) odpowiada za rozpoznawanie pojazdów poruszających się po drodze. Reguły mogą ograniczać liczbę zaznaczanych obiektów oraz oczekiwać detekcji obiektów na pasach ruchu znajdujących się w odpowiedniej relacji do samochodu będącego nośnikiem modułu. Oznaczone są samochody osobowe, ciężarówki, autobusy, traktory, itp. Przykładowe dane GT oznaczone na klatkach kamery dla tego modułu są pokazane na rysunku 2.7.

2.3.5. Znaki i sygnalizacja świetlna

Moduł percepcji TSR (*Traffic Sign Recognition*) odpowiada za rozpoznawanie oraz klasyfikowanie znaków drogowych. Odpowiada również za decydowanie o tym czy rozpoznany na zdjęciach znak dotyczy pasa ruchu, po którym porusza się pojazd. Moduł percepcji TLR (*Traffic Light Recognition*) zajmuje się rozpoznawaniem sygnalizacji świetlnej oraz jej stanu. Właściwie stanowi rozszerzenie modułu TSR. Interpretuje również strzałki warunkowe. Przykładowe dane GT tworzone dla tych modułów wyglądają jak na rysunku 2.8.

2.3.6. Linie, droga i wolna przestrzeń

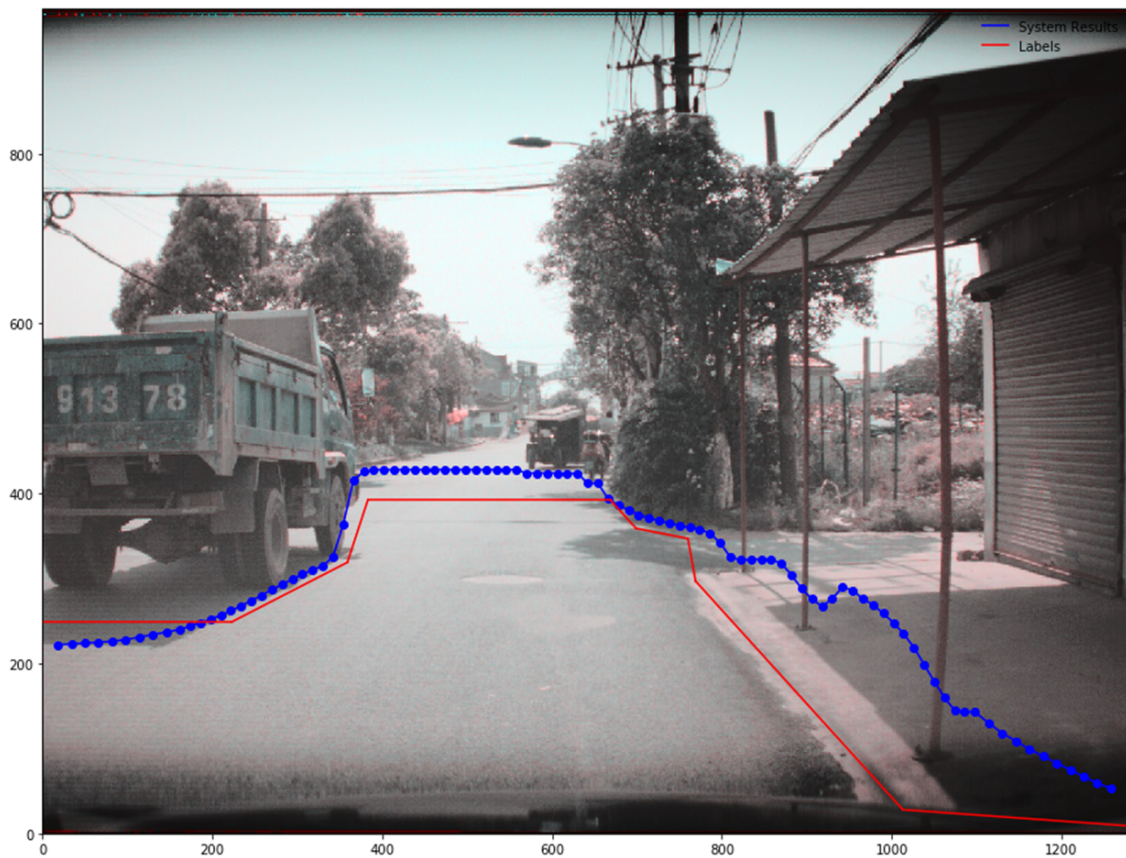
Osobny moduł percepcji odpowiada za oznaczanie linii oraz światła drogi. Klasyfikowany jest rodzaj linii (ciągła, przerywana). Zaznaczane są również linie dróg sąsiadujących z naszą, granice jezdni oraz pobocza (rysunek 2.9). Istnieje też moduł percepcji zajmujący się identyfikowaniem wolnej przestrzeni wokół samochodu (ang. *Free Space*). Również tej niekoniecznie będącej częścią regularnego ruchu ale mogącej być użyteczną w przypadku konieczności uniknięcia wypadku w sposób nie powodujący nowego niebezpieczeństwa. Dane GT oraz przykładowy wynik modułu percepcji są zaznaczone na rysunku 2.10.

2.4. Testowanie percepcji pojazdu - wyzwania

Metoda oceny jakości, zaproponowana w tej pracy, jest skoncentrowana na zastosowaniach w projekcie badawczo rozwojowym, a nie warunkach produkcyjnych. Powinna przede wszystkim być powtarzalna i niezależna od subiektywnej oceny inżyniera. Podczas testowania oprogramowania często korzysta się z testów regresyjnych badających czy kolejne wersje oprogramowania działają skuteczniej niż poprzednie oraz testów spraw-



Rysunek 2.9. Przykładowe dane GT tworzone dla linii na drodze



Rysunek 2.10. Odcinki wyznaczające dane referencyjne dla granic wolnej przestrzeni (czerwony) oraz punkty będące wynikiem przetwarzania obrazu przez moduł percepcji (niebieski)

dzających poprawność funkcjonowania w bardzo podstawowych aspektach (ang. *smoke test*). Testy te choć potrzebne, stanowią zaledwie małą część problematyki weryfikacji i walidacji autonomicznych rozwiązań, w szczególności testowania modułów percepcji.

2.4.1. Testowanie niedeterministyczne - dystrybucja danych

O testowaniu deterministycznym możemy mówić w sytuacji gdy wyniki testu są całkowicie zdeterminowane przez dane wejściowe i nie zmieniają się dopóki one nie ulegną zmianie. Podczas testowania oprogramowania wspierającego kierowcę sprowadza się to do wymagania, by metryki jakości wskazywały zawsze ten sam wynik w sytuacji gdy uruchamiamy moduł percepcji w tej samej wersji, na tych samych danych z sensora i w tym samym środowisku. Jednak aby udowodnić, że moduł percepcji spełni swoją funkcję w warunkach rzeczywistych nie można ograniczyć się do sprawdzenia go w jednej sytuacji, a raczej ustalić dystrybucję danych specyfikującą różne kombinacje warunków, w których analizujemy działanie systemów [1]. Taką dystrybucję danych realizuje się podczas jazd testowych, jednak o ile można mieć wpływ na porę dnia i szybkość testowego pojazdu, a nawet zaplanować ją tak by odbywała się w określonych warunkach pogodowych to jednak treść takich scenariuszy, a przez to i wartość zależy mocno od innych uczestników drogi, których zachowania nie można już przewidzieć. Należy założyć, że pojedynczy scenariusz uwzględniający konkretne trajektorie obiektów i warunki na drodze jest drobnym wycinkiem rzeczywistości – pojedynczą próbką z ogromnej populacji. Oznacza to, że testowanie wymaga metodologii uwzględniającej stochastyczną naturę weryfikowania skuteczności i skupiającej się na ocenie generalizacji wyników w szerokim kontekście. Kryterium akceptowalności poziomu wydajności powinno być zatem zdefiniowane na gruncie statystycznym. Znalezienie metody testowania zautomatyzowanych rozwiązań w samochodach o wysokim poziomie autonomiczności biorąc pod uwagę ważną rolę jaką odgrywa w nich sztuczna inteligencja oparta na sieciach neuronowych stanowi ogromne wyzwanie [53].

2.4.2. Przechowywanie i dostęp do danych

Wyzwania, przed którymi stają inżynierowie podczas testowania pojazdów autonomicznych są podobne do tych, jakie pojawiają się w każdej gałęzi przemysłu mającej do czynienia z dużymi zbiorami danych (ang. *Big Data*) [1]. Z pewnością jedną z trudności są wciąż niedopracowane standardy i brak szeroko uznanej metodologii koniecznej do analizy ogromnych ilości danych o złożonej strukturze. Gromadzenie danych jest procesem,

w którym nie można zaplanować i zoptymalizować wszystkich jego aspektów. Potencjalnych scenariuszy drogowych jest właściwie nieskończenie wiele (jeśli weźmiemy pod uwagę unikalnych uczestników ruchu) dlatego wybór spośród zdobytych danych po pierwsze tych poprawnych, a po drugie podzbioru tych rzeczywiście wartościowych jest niezwykle ważnym i trudnym zadaniem, a warto rozpocząć go już na poziomie zapisanych w samochodzie testowym danych. Również ze względu na przesył danych zebranych podczas dnia testowej jazdy w przeciągu ograniczonej ilości czasu, bezpiecznie i niezawodnie do centrum danych zajmującego się ich późniejszego przechowywaniem i dbającym o dostęp do nich w późniejszych fazach pracy z nimi. W związku z tym istotne jest wypracowanie sposobów zidentyfikowania typu informacji zawartej w danych oraz umiejętność oceny jej wartości. Jest to dodatkowo utrudnione przez ich złożoność. Złożoność ta jest przynajmniej częściowo spowodowana przez mocno zróżnicowane i liczne źródła generujące te dane (w przypadku samochodów autonomicznych są to różne rodzaje sensorów i odpowiadające im etykiety). Powiększające się wykładniczo zbiory danych wymagają coraz bardziej pojemnych i szybkich pamięci masowych, a ograniczona przepustowość łącz powoduje problemy z szybkim dostępem do nich. Koszty takich usług są znaczące [1]. Motywuje to do redukcji ilości przechowywanych danych, co w praktyce oznacza m.in. zidentyfikowanie i odrzucenie informacji nadmiarowych sklasyfikowanych jako nieistotne dla osiągnięcia celów projektu. Regularna analiza dużych zbiorów danych jest związana z utrzymywaniem i rozwijaniem potężnych klastrów obliczeniowych HPCC (ang. *High Performance Computing Cluster*) mogących przetwarzać tak duże zbiory i będących pośrednikami między centrum danych, a inżynierami pracującymi nad nimi. Podsumowując, rozwój technologii pozwalającej na gromadzenie i przetwarzanie dużych ilości danych musi iść w parze z rozwojem metod analizy uwzględniającej charakterystyczne własności danych oraz ich wielką skalę aby umożliwić nie tylko skuteczność ale też sprawny przebieg procesu testowania i rozwoju inteligentnej percepcji oraz optymalizację jego kosztów [11]

2.4.3. Etykietowanie danych

Aby podjąć się ewaluacji jakości wyników modułów percepcji konieczne jest uprzednie etykietowanie danych z sensorów tak, aby uzyskać dane referencyjne takie, jakich przykłady pokazano w poprzedniej sekcji. Etykietowanie obiektów w danych wizyjnych polega, w przybliżeniu, na wyznaczeniu granic prostokątnego regionu rozpoznania i sklasyfikowaniu go w specjalnym programie do tworzenia etykiet. Programy takie mogą sugerować etykiety same je generując (mówi się wtedy o etykietowaniu półautomatycznym,

wspieranym przez sztuczną inteligencję [54]), jednak praca osoby przygotowującej dane referencyjne wciąż będzie polegać na weryfikacji tych sugestii i ewentualnych korektach. Jeżeli w jednej klatce filmu mamy od kilku do kilkudziesięciu obiektów z różnych klas, klatkaż filmu to 30 FPS, a przykładowa sekwencja scen może trwać od kilku do kilkunastu minut, to nietrudno sobie wyobrazić jak czasochłonnym i żmudnym jest takie zadanie. Zakładając, że takie sceny są wybierane spośród danych zebranych z floty kilkudziesięciu samochodów, nagrywających przez kilkanaście godzin dziennie, przez kilka tygodni, trzeba sobie uzmysłowić jak wiele danych należy opracować, nawet jeśli wybrane do tego zostaną tylko niektóre sceny. Dodatkowo niemożliwym jest, by człowiek skupiał się nad tego typu czynnością przez dłuższy okres czasu. Z tych powodów w pracę nad tworzeniem danych GT zaangażowane są setki ludzi.

2.4.4. Resymulacja

Aby uniknąć przeprowadzania jazd testowych w celu weryfikacji każdej nowej wersji oprogramowania, tak jak wspomniano powyżej, nagrywa się je, aby później wykorzystać je ponownie w sposób wirtualny. Zapisane dane z sensorów, aby zostać wykorzystane w procesie uczenia, testowania lub walidacji, muszą zostać odtworzone i w sposób uporządkowany dostarczone na wejście modułów percepcji tak, aby uzyskać ich wyniki, które następnie można poddać procesowi ewaluacji z wykorzystaniem odpowiadających im danych referencyjnych. Taki sposób testowania nowych wersji modułów percepcji nazywa się resymulacją - ReSim (ang. *re-simulation*). Resymulacja danych, w celu uzyskania detekcji, nie musi się odbywać w samochodzie. Efektywną metodą resymulacji, w przypadku gdy resymulowane mają być setki scen, jest uruchomienie jej na HPCC. Resymulacja jednej minuty sceny testowej i pozyskanie detekcji z wykorzystaniem jednego rdzenia może trwać nawet 10 minut. Posiadając 100 scen testowych trwających 3 minuty każda i HPCC dysponujące 50 rdzeniami można uzyskać wyniki detekcji dla pięciu godzin nagranych danych w jedną godzinę. Im więcej rdzeni tym więcej scen może być resymulowanych jednocześnie. W przypadku danych uzyskanych z kamery wyniki detekcji poszczególnych obiektów w kolejnych klatkach trwania filmu są zapisywane do osobnego pliku, który później zostanie porównany z etykietami stworzonymi manualnie, co pozwoli na ocenę jakości działania modułu percepcji. w przypadku testowania modułów percepcji na tym kończy się procedura, co sprawia, że jest to pętla otwarta. Aby umożliwić testowanie funkcji pojazdu, związanych, na przykład, z aktywnym bezpieczeństwem, pojazd musi wejść w interakcję ze środowiskiem, co nie jest możliwe w przypadku użycia nagranych danych. Aby domknąć pętlę i umożliwić to, można posłużyć się środowiskiem wirtualnym, ponieważ

istnieje możliwość modyfikowania go w zależności od zachowania pojazdu na skutek decyzji algorytmów. Choć obecnie wirtualne środowisko symulatora nie gwarantuje jeszcze wizualizacji otoczenia na poziomie dokładności dorównującym danym pochodzącym z jazd testowych, to rozwój technologii symulacji pozwala mieć uzasadnioną nadzieję, że wkrótce będzie to możliwe. Już jednak obecnie dane syntetyczne pozwalają rozszerzyć możliwości testowania systemów aktywnego bezpieczeństwa na obszary praktycznie nieosiągalne przy klasycznym podejściu, na przykład przez symulację sceny, w której kolizja lub inne niebezpieczne zdarzenie jest nieuniknione. Pozwala to rozszerzyć zbiór sytuacji testowych o takie, których nie można zarejestrować podczas zbierania danych eksperymentalnych.

Rozdział 3

Ewaluacja interpretacji otoczenia

3.1. Prostokątne regiony rozpoznania

Dane związane z testowaniem modelu modułów percepcji można podzielić ze względu na format danych jakich one używają aby opisać elementy obecne w scenie testowej w sposób uproszczony i zrozumiały dla komputerowych algorytmów. Za opis najważniejszych obiektów znajdujących się na drodze odpowiadają moduły percepcji opierające się między innymi na systemach wizyjnych i sieciach neuronowych, które wykorzystują prostokątne regiony rozpoznania (ang. *bounding boxes*) do zaznaczania rozpoznanych obiektów w każdej oddzielnej klatce strumienia wideo. Przy założeniu, że wszystkie prostokąty mają boki równoległe do krawędzi ramki wideo, mogą być one przechowywane jako cztery liczby. Współrzędne początku i końca przekątnej lub jako współrzędne środka oraz długość i szerokość. Chociaż w tej pracy przykłady dotyczące danych wideo zawierają tylko pola w tej formie, założenie o równoległości boków prostokątów do krawędzi ramki wideo nie ogranicza stosowalności opisanego metodologii do bardziej ogólnych sytuacji, w których obiekty posiadają też przypisany kąt rotacji względem samochodu będącego nośnikiem inteligentnych sensorów (patrz sekcja 4.4.2). Prostokątne regiony ograniczające służą do oznaczania obiektów takich jak piesi, inne pojazdy, ich światła (oddzielnie), znaki drogowe, progi zwalniające oraz sygnalizacja świetlna. Dane odniesienia, do których można porównywać wyniki pochodzące z modułów i czujników są tworzone ręcznie dla każdej klatki i mają postać prostokątów wzorcowych (dane referencyjne), które pokrywają obiekty zainteresowania dla danej funkcji. Należy ocenić, jak dobrze detekcje proponowane przez systemy rozpoznawania obiektów pasują do każdego obiektu oznaczonego w danych referencyjnych. A zatem zadanie u podstaw polega na stworzeniu metodologii porównania prostokątów.



Rysunek 3.1. Prostokątne regiony rozpoznania oznaczające pieszych, samochody, znaki i sygnalizację świetlną

3.1.1. Inne klasy danych

Linie na jezdni z reguły są oznaczane przy pomocy jednej lub kilku krzywych określonych wielomianem stopnia co najwyżej trzeciego oraz punktami wyznaczającymi jego wycinek. Oznacza to, że aby przechować jeden segment linii zapisujemy sześć parametrów: a' , a'' , a_0 , a_1 , a_2 , a_3 reprezentujących wielomian ograniczony:

$$P(x) = a_3x^3 + a_2x^2 + a_1x^1 + a_0, \quad x \in [a', a'']. \quad (3.1)$$

W ten sposób oznaczane są granice pasa na którym porusza się samochód będący nośnikiem modułów percepcji, jego oś ruchu, pasy sąsiednie, linie krawędziowe (granica drogi), pobocza. Przykład widoczny na rysunku 2.9. Punktami oznaczane są inne elementy otoczenia (np. elementy odbłaskowe umieszczone na jezdni) lub granica wolnej przestrzeni wokół samochodu widoczna na rysunku 2.10.

3.2. Wyznaczanie jakości modułu percepcji

Kluczowe wskaźniki efektywności czyli KPI (*ang. Key Performance Indicator*) są to współczynniki pozwalające w sposób mierzalny analizować procesy i stopnie realizacji celów organizacji. Mają one przypisaną wartość liczbową, ale ich metoda obliczania oraz sposób interpretacji są uzależnione nie tylko od postawionego celu ale również od branży, wielkości przedsięwzięcia, strategii oraz specyficznych cech zadania. Dobrze skonstruowane KPI powinny umożliwiać szybkie podejmowanie decyzji, planowanie i priorytetyzację działań w szczególności w ramach reagowania na pojawiające się problemy. Wiążą się one bezpośrednio z optymalizacją wykorzystywania posiadanych zasobów w firmie. Powinny być obiektywną informacją, która pozwala nawet osobie o stosunkowo niskiej znajomości szczegółów analizowanego procesu celnie ocenić wyniki oraz jakość podejmowanych działań. Generalne sposoby projektowania KPI określa norma ISO 22400 [55, 56].

W kontekście analizy jakości modułów percepcji w samochodzie autonomicznym KPI będą opisywać jakość rozpoznania i śledzenia obiektów znajdujących się w otoczeniu samochodu w odniesieniu do opisu przygotowanego uprzednio manualnie lub półautomatycznie. Analizowane dane to zapisy dotyczące surowych danych otrzymywanych z sensorów, dane wstępnie przetworzone przez algorytmy percepcji, metadane uzyskane w trakcie zbierania danych oraz przede wszystkim wyniki działania sieci neuronowych rozpoznających oraz kategoryzujących obiekty w otoczeniu pojazdu. W efekcie otrzymujemy złożone zbiory danych opisujące czasoprzestrzenne, statyczne i dynamiczne właściwości elementarnych obiektów składających się na wycinek rzeczywistości, z którym autonomiczny samochód musi wejść w interakcję lub na podstawie, którego ma podjąć decyzję o swoim zachowaniu. Złożenie tak ogromnego zbioru danych w spójną całość jest procesem wieloetapowym, dlatego obliczanie jakości rozpoznania obiektów i rekonstrukcji całego otoczenia oraz jego właściwości są również procesami wieloetapowymi. Na każdym poziomie łączą one potrzebę uogólniania opisu jakości z koniecznością przekazania informacji kluczowych ze względu na bezpieczeństwo pojazdu.

Przez analizę jakości zbiorów danych przygotowanych do testowania modułów percepcji oraz algorytmów sterowania będziemy rozumieć analizę wszystkich danych zebranych podczas wielu jazd testowych rozumianych jako zbiór wspomnianych wycinków rzeczywistości - scen. Pobierane są one w różnych warunkach pogodowych, przy zmiennym poziomie oświetlenia, natężenia ruchu na drodze i w otoczeniu różniącym się gęstością i wysokością zabudowy, złożonością struktury drogowej i liczby potencjalnych obiektów takich jak piesi, znaki i sygnalizacje świetlne. Analiza przeprowadzona na takim zestawie

danych również korzysta z opisu obiektów wspomnianego wyżej, ale przenosi się na wyższy poziom starając się ocenić i formalnie opisać różnorodność zestawów scen rozumianych jako zbiory trenujące, testujące i walidujące algorytmy i sieci neuronowe wykrywające obiekty i procesy podejmujące decyzje o sterowaniu pojazdem.

3.2.1. Ocena jakości detekcji

Ewaluacja na poziomie klas danych opisanych powyżej będzie oznaczać konieczność definicji odpowiednich miar podobieństwa. Aby rozpocząć te rozważania warto przypomnieć matematyczną definicję tego pojęcia. Dla zbioru \mathcal{R} , *miarą podobieństwa* będziemy nazywać funkcję $s : \mathcal{R} \times \mathcal{R} \rightarrow \{p \in \mathbb{R} : 0 \leq p \leq 1\}$ jeśli spełnione są następujące dwa warunki:

$$\forall x, y \in \mathcal{R} : s(x, y) \leq s(x, x) \quad (3.2)$$

oraz

$$\forall x, y \in \mathcal{R} : s(x, y) = s(y, x). \quad (3.3)$$

Warto zauważyć, że trzecia własność:

$$\forall x, y \in \mathcal{R} : s(x, y) = 1 \Leftrightarrow x = y \quad (3.4)$$

pomimo, iż nie jest formalnie wymagana, jest niezwykle wartościowa. Szczególnie w rozpatrywanym w pracy przypadku ze względu na to, że na bazie utworzonych miar podobieństwa utworzone zostaną kolejne, bardziej ogólne miary. W praktyce, której dotyczy ta praca, o użyteczności takiej funkcji decyduje empirycznie potwierdzona zdolność do wyraźnego oddzielania detekcji poprawnych od wadliwych.

3.2.2. Indeks Jaccarda

W kontekście porównywania prostokątnych regionów rozpoznania istnieje wiele miar podobieństwa i metryk opisujących jakość reprezentacji manualnie przygotowanych prostokątów referencyjnych przez prostokąty pochodzące z modułów percepcji (lub innych sieci neuronowych) [57], [58]. Najczęściej wykorzystywanym konstruktem posiadającym wiele cech sprawiających, że sprawdza się on jako podstawowe narzędzie do analizy jakości w tym zadaniu jest *Indeks Jaccarda* [59]. Nazywany też IoU (ang. *Intersection over*

Union). Jest to funkcja $\mathcal{J} : \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^n) \rightarrow [0, 1]$, $n \in \mathbb{N}$ zdefiniowana następująco:

$$\mathcal{J}(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (3.5)$$

Indeks Jaccarda jest miarą podobieństwa spełniającą wszystkie trzy warunki z poprzedniej podsekcji. Jest szeroko stosowany w zadaniach ewaluacji segmentaryzacji obrazu [60]. W zadaniach takiego rodzaju stanowi dobre rozwiązanie ponieważ jest to generalne narzędzie do porównywania dwóch nieregularnych obiektów. Można wskazać bardziej odpowiednie podejście [61], ale trudno wskazać bardziej naturalne podejście. Obliczenie indeksu Jaccarda polega na podzieleniu liczby poprawnie pokrytych pikseli przez sumę wszystkich pikseli biorących udział w porównaniu. Problem ewaluacji prostokątnych regionów rozpoznania jest szczególnym przykładem ewaluacji segmentacji obrazu i powinien być potraktowany osobno. Ocena jakości oparta o indeks Jaccarda formalnie zadziała, ale nie zostanie wykorzystana informacja o tym, że porównywane obiekty zawsze należą do tej samej klasy posiadającej dobrze znane geometryczne właściwości. Podobną krytykę tego podejścia można znaleźć w [62], gdzie ponadto odnajdziemy również alternatywne podejście do konstruowania nowych miar na bazie właściwości prostokątów.

3.2.3. Miary jakości detekcji oraz ich zastosowania

Większość prac dotyczących lub zawierających analizy danych w postaci prostokątów opiera się na indeksie Jaccarda lub podobnych miarach, służących do oceny detekcji pochodzących z przetwarzania danych wideo. Niektóre prace prezentują porównania wielu różnych miar jakości, ale wciąż nie ma zgody co do tego, które stanowią najlepszy wybór gdy chcemy oceniać wyniki eksperymentów, ponieważ większość technik ewaluacji opisanych w literaturze nie odnosi się do złożoności i zakresu zagadnień, które leżą u podstaw projektowania dobrej metodologii ewaluacji [63]. Z tego powodu stale opracowywane są nowe metodologie obiektywnej analizy jakości wykrywania i śledzenia obiektów, ale większość z nich nadal wykorzystuje jako podstawę indeks Jaccarda [64], [65], [66], [67]. W ramach zadań związanych z rozwojem detektorów powstało wiele standaryzowanych zbiorów danych, na których można dokonać ewaluacji działania algorytmu. Zbiory te zawierają również: adnotacje stanowiące referencje dla obiektów, dla których oczekujemy, że zostaną rozpoznane i wytyczne opisujące zasady tworzenia takich adnotacji. Wszystko to w celu ujednoczenia testów i umożliwienia porównania algorytmów detekcji i śledzenia. Jednak cel ten pozostaje niemożliwy do osiągnięcia, jeżeli stosuje się różne metryki oceny [64]. Popularnym zastosowaniem miary podobieństwa jest traktowanie jej jako

funkcji straty przy trenowaniu algorytmów śledzących. Praca [68] przedstawia metodę optymalizacji średniej straty IoU w sieciach neuronowych. Wykazano, że strata działa lepiej korzystając z miary Jaccarda niż tradycyjnie stosowana strata entropii krzyżowej. Konieczność zbudowania metodyki ewaluacji motywowana jest najczęściej różnymi problemami dotyczącymi konkretnych algorytmów detekcji. Algorytm śledzenia przesunięcia średniego z funkcją przestrzenno-kolorową [69] i parametryzowany filtr Kalmana użyty w celu śledzeniu obiektów na wideo [70] to przykłady algorytmów śledzących, które wymagają oceny. Artykuł [71] opisuje protokół oceny jakości detekcji i śledzenia twarzy, tekstu i pojazdów na wideo. Praca [72] opisująca tak zwane „wyzwanie DAVIS”, choć skupia się na segmentacji obrazu, również proponuje metodologię oceny do opisu jakości wyników algorytmów w analizowanej klatce filmu oraz sumaryczną jakość w całej scenie. Autorzy w [73] przedstawili połączone kryterium oceny wydajności, biorąc pod uwagę wiarygodność, przetwarzanie w czasie rzeczywistym i implementację na sprzęcie algorytmu śledzenia. W [74] można znaleźć algorytm, który wyprowadza reguły rozmyte w celu scalenia propozycji prostokątnych regionów rozpoznania w jeden reprezentacyjny, który opisuje unikatowy obiekt. Autorzy podkreślają potrzebę stworzenia narzędzia opisującego relacje dwóch prostokątów poprzez ich geometryczne cechy, spójność ruchu i podobieństwo wyglądu. Wykrywanie i śledzenie obiektów w realistycznych scenach to złożony problem i pozostaje aktywnym obszarem badań dotyczących wizji komputerowej [75]. Potrzeba odpowiedniej metodologii analizy jest zauważalna zarówno w zagadnieniach ewaluacji śledzenia obiektów, jak i oceny jakości fuzji danych z różnych sensorów. Dlatego [76] definiuje metodologię oceny, która może być zastosowana również w tej dziedzinie. Zaproponowane przez autora metryki „kontekstowe” w założeniu mają wspierać standardowe metryki, aby radzić sobie z różnymi wyzwaniami, które stoją na drodze do prawidłowej oceny algorytmów śledzenia/fuzji.

Indeks Jaccarda wydaje się pod wieloma względami przewyższać istniejące rozwiązania i jest szeroko stosowany w różnych projektach, które wymagają miernika oceny. W [77] autorzy używają go do badania wpływu małych błędów w danych uczących na wyniki wytrenowanego algorytmu śledzenia. Warto jednak zwrócić uwagę, że w tej pracy zostaje również postawiony nacisk na to, że w podobnych zadaniach bardzo ważna jest precyzja i możliwość dekompozycji informacji. Są prace, których autorzy dostrzegają braki tej metryki i proponują jej rozszerzenie [78] lub [79], są to prace motywowane analizą straty IoU stosowanej podczas projektowania i konfigurowania modelu sieci przy użyciu stochastycznego zejścia gradientowego [80]. Wykorzystanie indeksu Jaccarda do porównywania ramek ograniczających wiąże się z ryzykiem pominięcia wielu informacji na podstawowym

poziomie analizy danych. Istnieje potrzeba i możliwość stworzenia bardziej szczegółowej miary niż IoU, co podkreślone zostało w pracach [62], [81]. Pożądana miara powinna nie tylko doprecyzować ocenę, ale również posiadać te same właściwości, które sprawiają, że Indeks Jaccarda jest tak szeroko stosowany. Więcej informacji o własnościach miary porównawczej znajduje się w sekcji 4.5.1.

3.2.4. Detekcja prawdziwie pozytywna

Powyżej opisane przykłady miar jakości z indeksem Jaccarda na czele stanowią podstawę do definiowania tak zwanego rozpoznania prawdziwie pozytywnego - TP (ang. *True Positive*). Jest to w ujęciu lokalnym (czyli dla pojedynczego porównania prostokątów w jednej klatce filmu) sytuacja, w której prostokąt referencyjny uznaje się za poprawnie reprezentowany przez prostokąt będący wynikiem działania sieci neuronowej na podstawie satysfakcjonująco wysokiej wartości metryki jakości. Przykładowo $IoU > 50\%$ stanowi podstawę takiej klasyfikacji według zasad ustalonych w [82]. Sytuacja gdy prostokąt referencyjny, który nie został oznaczony przez sieć według takiego rozgraniczenia jest oznaczana jako rozpoznanie fałszywie negatywne FN (ang. *False Negative*). Ostatecznie prostokąty pochodzące z sieci i nie stanowiące poprawnej detekcji dla żadnego z prostokątów opisanych w referencji są oznaczane jako rozpoznania fałszywie pozytywne (ang. *False Positive*). Bazując na takiej klasyfikacji detekcji, możemy obliczyć dwie istotne metryki mówiące o jakości detektora czyli precyzja oraz czułość, gdzie

$$\begin{aligned} \text{Precyzja} &= \frac{\text{liczba } TP}{\text{liczba } TP + \text{liczba } FP}, \\ \text{Czułość} &= \frac{\text{liczba } TP}{\text{liczba } TP + \text{liczba } FN}. \end{aligned} \quad (3.6)$$

Aby podsumować obie te metryki można wykorzystać ich ważoną średnią harmoniczną, czyli miarę

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{Precyzja} \cdot \text{Czułość}}{\beta^2 \cdot \text{Precyzja} + \text{Czułość}}, \quad (3.7)$$

gdzie parametr β pozwala na ustalenie względnej istotności wpływów precyzji i czułości na ostateczną wartość tej metryki [63]. Dokładniej o tym jak wygląda zależność między tymi współczynnikami detektora informuje nas współczynnik AP (ang. *Average Precision*) [82]. Należy zauważyć, że zwykle detektor bardziej precyzyjny będzie posiadał więcej rozpoznań fałszywie negatywnych, a przez to będzie mniej czuły. Z drugiej strony detektor wysoce czuły ma większe ryzyko wygenerowania obserwacji fałszywie pozytywnej, a

przez to tracić precyzję. Identyczny efekt na te metryki ma dobór metodologii definicji rozpoznania prawdziwie pozytywnego co sprawia, że jej wiarygodność ma bezpośrednie przełożenie na wiarygodność zdefiniowanych powyżej metryk. Dodatkowo obliczanie tych metryk poprzez sumowanie ilości TP, FN i FP ze wszystkich klatek filmu ignoruje fakt, że poszczególne detekcje dotyczą często tych samych obiektów pojawiających się w zasięgu sensorów przez dłuższy okres czasu. Z tego powodu nie niosą czytelnej informacji na przykład o tym czy detekcja pojawiła się tuż po wkroczeniu obiektu w zasięg sensora i straciła go w późniejszym okresie, czy odwrotnie.

3.2.5. Podsumowanie jakości w czasie

Jeśli chodzi o naturalnie związany z lokalną jakością w klatce problem podsumowania oceny sekwencji prostokątów opisujących jeden spójny obiekt w zestawie kolejnych klatek wideo poruszony na przykład w [83], istnieją inne podejścia niż najbardziej klasyczne polegające na policzeniu średniej jakości (zastosowane w wyzwaniu DAVIS). Przykładami są odmiany metryki OSPA [84], [85], [86], którą cechuje możliwość uszczegółowienia informacji, ponieważ miara ta składa się z trzech komponentów, z których każdy osobno uwzględnia lokalizację, częstość i błędy etykietowania. Autorzy w [87] proponują nową prostą metodologię oceny algorytmów śledzących, która jako cel stawia sobie możliwość łatwego do interpretacji porównania algorytmów śledzenia (ang. *tracker*). Zaproponowana metodologia uznaje równoważność algorytmów śledzących na podstawie analizy istotności statystycznej rozbieżności i praktycznych różnic. Metrykami oceny obiektów w czasie są: podobieństwo korelacji ważonej stanów [88], trudność śledzenia [89], metryka oparta na teorii informacji [90], miara asocjacji kształtu bazująca na odległości Minkowskiego [91] i inne [92], [93], [94]. Niestety żadna z nich nie stawia wystarczającego nacisku na szybkie inicjowanie rozpoznania nowych obiektów, co jest kluczowe w sytuacjach dynamicznej jazdy, które wymagają natychmiastowej reakcji.

3.2.6. Wpływ dokładności detekcji na układ sterowania

Brak wykrycia samochodu na drodze oznacza, że nie zostanie on uwzględniony przez algorytmy aktywnego bezpieczeństwa, co w oczywisty sposób stwarza możliwość kolizji pojazdów i zagraża bezpieczeństwu pasażerów. Brak detekcji pieszego na przejściu dla pieszych może skutkować potrąceniem go. Brak klasyfikacji znaku dotyczącego obowiązującego ograniczenia prędkości może oznaczać utrzymywanie nieoptymalnej lub nadmiernej prędkości powodującej zagrożenie. Jednak detekcja nieperfekcyjnie oddająca właściwości

obiektu może mieć podobnie negatywne skutki co jej brak. Skuteczna i bezpieczna jazda wymaga korygowania kierunku i prędkości pojazdu, które uwzględniają aktualną sytuację na drodze oraz sytuację przewidywaną w najbliższej przyszłości [95]. Algorytmy sterujące pojazdem, np. adaptacyjny tempomat, aby działać poprawnie wymagają odpowiedniego poziomu pewności odnoszącego się do informacji o otoczeniu pojazdu [96]. Każdy sensor posiada swoje zalety i wady [97], [98]. Dlatego szczegółowa wiedza o wiarygodności rozpoznawanych przez nie obiektów jest kluczowa dla bezpieczeństwa pojazdu.

3.3. Wybór reprezentatywnego zbioru danych

Podczas projektowania i rozwoju autonomicznych systemów wspierania kierowcy kluczową rolę odgrywają aspekty związane z bezpieczeństwem. Przetestowanie protokołów sterujących pojazdem, a w szczególności jego percepcji we wszystkich możliwych sytuacjach jest w rzeczywistości niemożliwe ponieważ potencjalnych scenariuszy drogowych jest zdecydowanie zbyt wiele [99]. Podobnie jest w przypadku ewaluacji komponentów odpowiedzialnych za interpretację środowiska. Proces wyboru skończonego zbioru scenariuszy testowych w obu tych przypadkach stanowi wyzwanie i aktualnie często sprowadza się do inżynierskiej oceny eksperta zajmującego się danym zagadnieniem. Jest to „wąskie gardło” procesu ewaluacji ponieważ weryfikacja skomplikowanych sieci neuronowych pracujących w złożonym środowisku jakim są rzeczywiste wyzwania na drogach wymaga dużych i różnorodnych zbiorów danych, aby zapewnić odpowiedni poziom bezpieczeństwa [13]. Ze względu na przeznaczenie można wydzielić trzy rodzaje zbiorów scenariuszy użytecznych w rozwoju algorytmów samochodów autonomicznych w tym inteligentnej percepcji:

- **Zbiór danych uczących** – służy do uzyskiwania i trenowania parametrów sieci neuronowej.
- **Zbiór danych testowych** – służy do weryfikacji efektów uczenia sieci.
- **Zbiór danych walidacyjnych** – służy do oceny skuteczności algorytmów.

Obecnie przez duże zbiory danych mamy na myśli kolekcje zawierające kilkadziesiąt petabajtów (PB) danych. W ostatnich latach wymagania co do rozmiarów danych na jakich należy testować sieci znacznie się zwiększyły, a jednym ze sposobów by sobie z tym poradzić jest rozwój technologii umożliwiającej testowanie algorytmów samochodów autonomicznych z wykorzystaniem symulatorów (rysunek 3.2). W ten sposób można elastycznie



Rysunek 3.2. Przykłady obrazu otoczenia syntezywanego przez symulator jazdy CARLA [100], który jest uznawany za najlepszy spośród otwartych symulatorów przeznaczonych do rozwoju systemów autonomicznej jazdy [101]

uzupełniać zbiory rzeczywistych scenariuszy nagranych podczas jazd testowych i wykorzystać je do poprawienia procesu nauki, testowania funkcji i walidacji. W szczególności w sposób syntetyczny można odtworzyć sytuacje, których pojawienia nie można zaplanować podczas jazd testowych. Przykładami są scenariusze gdzie kolizja pojazdów lub podobnie niebezpieczna sytuacja na drodze jest nieunikniona. Tego typu dane pozwalają w sposób pełniejszy przeanalizować zachowanie systemów pojazdu w skrajnych przypadkach.

3.3.1. Charakterystyka dużych zbiorów danych

Jak wspomniano w poprzednich sekcjach bardzo ważnym wyzwaniem podczas testowania percepcji pojazdu jest zdefiniowanie KPI pozwalających porównywać jakość różnych algorytmów pracujących na zadanym zbiorze danych. Podstawowym problemem w tym procesie jest to, że mając dwa różne zbiory treningowe, możliwym jest wytrenowanie sieci w istotnie różny sposób [102]. Dokonując walidacji systemu na dwóch różnych zbiorach zwykle uzyskujemy dwa różne zestawy metryk opisujących jakość działania systemów. Ta zależność generuje potrzebę opracowania metodologii formalnego opisu scenariuszy testowych oraz propozycji miar służących do porównywania ze sobą dużych zbiorów scenariuszy. Opracowane kryteria porównawcze powinny umożliwić między innymi wiarygodne porównywanie wyników pomiaru jakości, dzięki możliwości zdefiniowania wymagań dotyczących podobieństwa zbiorów danych, na których dwa różne algorytmy są weryfikowane. Formalny opis zbiorów danych może bazować w dużej mierze na opisie różnych rodzajów metadanych, które je charakteryzują. Metadane odgrywają kluczową rolę w procesie meta-uczenia [103], czyli dziedzinie badań mających na celu poprawę wydajności uczenia sieci neuronowych. Metadane w przypadku zestawów danych motoryzacyjnych mogą być

traktowane jako uzupełnienie surowych danych rejestrowanych bezpośrednio z czujników pojazdu (np. pozycja GPS, prędkość, kąta odchylenia pionowego, warunki pogodowe itp.) lub dodawane później w wyniku przetwarzania surowych sygnałów przez funkcje systemu (np. wysokość słońca, liczba pieszych w klatce itp.) [104]. W [105] autorzy zbadali podobieństwa między zbiorami danych i opisano dwa podejścia: klastrowanie zbiorów danych przy użyciu korelacji błędów oraz przy użyciu rang. Natomiast w [106] do scharakteryzowania zbiorów danych zaproponowano wykorzystanie miar złożoności.

3.3.2. Wizualizacja złożonych danych

Jednym z użytecznych sposobów opisu i eksploracji dużych zbiorów danych jest wizualizacja. W literaturze ta tematyka jest często poruszana ponieważ dobra wizualizacja pozwala na skuteczne wyodrębnienie kluczowych informacji z dużych chmur danych. W [107] autorzy przedstawiają narzędzie o nazwie "Time-tunnel" do wizualizacji wielowymiarowych szeregów czasowych za pomocą wykresów 3D. Podejście umożliwia eksplorację i wygodne porównywanie wzorców, aby znaleźć silny związek między nimi. Analizując prace podejmujące temat dużych zbiorów danych dotyczących motoryzacji można dostrzec, że dane z czujników samochodowych postrzegane są przez autorów jako czasoprzestrzenne, co umożliwia automatyczną analizę i wizualizację geoprzestrzenną [108]. Inny przykład wizualizacji dużych zbiorów danych pochodzi z pracy [109], która opisuje jak wykorzystać geometrię współrzędnych równoległych do prezentacji danych wielowymiarowych i jak na tej podstawie prowadzić wizualną detekcję wzorców. Autor stwierdza, że złożonych danych nie można wyczerpująco opisać za pomocą jednej wizualizacji i dopiero zestawienie wielu różnych jej typów ujawnia różne możliwe sposoby opisu tych samych danych i jest kluczem do ich głębokiego zrozumienia. Aby wzmocnić efekt, dobrze jest opracować sposób koordynacji tak wielu punktów widzenia. Po różne typy wizualizacji danych wielowymiarowych sięgają autorzy [110]. Praca opisuje jak na podstawie macierzy korelacji zdecydować, które dwie zmienne należy wybrać, aby zbadać ich rozkład danych. Wybrane cechy mogą być reprezentowane jako trójwymiarowe sfery reprezentujące klastry, ich promień jest proporcjonalny do wielkości klastra, a kolor dobrany na podstawie zmiennej nieuwzględnionej podczas klastrowania. Artykuł [111] opisuje metodę wizualizacji przestrzennie ustalonych wielowymiarowych danych wolumetrycznych, która cechuje się zachowaniem właściwości przezroczystości przy jednoczesnym zapewnieniu łatwych do interpretacji wartości skalarnych oznaczonych kolorami. Autorzy znajdują kompromis między wzajemnie sprzecznymi celami dzięki redystrybucji nieprzezroczystości woksela przez uczynienie jego fragmentu bardziej nieprzezroczystym (czyniąc inne części bardziej

przezroczystymi). Można stwierdzić, że choć istnieją sposoby opisywania danych wielowymiarowych i czasoprzestrzennych charakterystyczne dla danych dotyczących motoryzacji, a związanych z pojedynczymi scenariuszami testowymi, to jednak istnieje wyraźna luka albo brak metodologii umożliwiającej wyodrębnienie z danych kluczowych informacji. Chodzi o informacje, które są ważne z punktu widzenia uczenia się i testowania samochodowych detektorów i algorytmów śledzących. Pożądana metodyka powinna obejmować nie tylko ilość danych, ale także ich jakość i zupełność, uwzględniając różnorodność całych zestawów złożonych ze zbiorów wielowymiarowych informacji [99].

3.3.3. Analiza trajektorii

Śledzenie obiektów definiuje się jako problem wyodrębniania ruchu obiektu z sekwencji klatek filmu albo skanów radarowych lub lidarowych. Następnie dokonuje się szacowania jego trajektorii czyli śladu jego podróży przez obszar znajdujący się w zasięgu sensorów. W literaturze szeroko omawiany jest temat analizy trajektorii ruchu. Jedno z najlepszych podejść jest przedstawione w [112] i nazywa się *TrajAlign*. Metoda ta stanowi sposób wyrównywania trajektorii umożliwiający wiarygodne określenie ich podobieństwa. Polega on na wykorzystaniu macierzy odległości oraz algorytmu programowania dynamicznego Needlemana i Wunscha [113]. W pracy [114] autorzy optymalizują obliczanie odległości trajektorii metodą dynamicznie skalującej się transformacji i używają pól wektorowych na różnaitości zamiast odległości euklidesowej. Praca [115] skupia się na wygładzaniu trajektorii przez klastrowanie z wykorzystaniem odległości Eulera i dwuwymiarowej funkcji będącej falką na otrzymanych grupach. Zrekonstruowaną trajektorię można traktować jako odsumione dane. Autorzy [116] w celu zdefiniowania nowatorskiego pomiaru podobieństwa, który uwzględni równoległe cechy czasoprzestrzenne i semantyczne, wykorzystują wielowymiarową macierz trajektorii czasoprzestrzennej i rozkład według wartości osobliwych (SVD) do wydobycia podstawowych cech trajektorii. Czasoprzestrzenne kojarzenie trajektorii jest również rozważane w [117]. Praca [118] zawiera przegląd istniejących podejść do funkcji podobieństwa trajektorii i ocenia je pod względem kosztów obliczeniowych i pamięci. W pracy [119] autorzy pokazują metodę normalizacji pozycji dla procesu dopasowania trajektorii uniezależniającą ją od translacji, rotacji i skali. Praca [120] przedstawia metodę obliczania podobieństwa między trajektoriami opartą na współczynniku korelacji Pearsona i współczynniku determinacji. W [121] autorzy zaproponowali miarę podobieństwa trajektorii opartą na najdłuższym wspólnym podciągu (ang. *Longest Common Subsequence* – LCS) i przetestowali ją na bazie danych zawierającej zapisy po-

ruszających się po mapie miasta pojazdów. Artykuł [122] przedstawia odporną na szum wersję metody LCS.

Wszystkie te metryki mogą być przydatne do analizy trajektorii. Jednak celem analizy trajektorii przedstawionej w tej pracy jest zaproponowanie metryki porównawczej trajektorii, która uwzględnia pozycję na siatce zajętości rozpiętej wokół poruszającego się pojazdu i czas, w którym dany obiekt zajmował obszar na siatce. Żadna z wymienionych wyżej metod nie umożliwia wprost porównywania układów trajektorii, co zapewniłoby funkcjonalność porównania całych scenariuszy, a konsekwencji ich zbiorów.

3.3.4. Metryka Wassersteina

Zaproponowanie metryki operującej na wielowymiarowych danych, które jak wspomniano są często trudne do wizualizacji, jest niełatwym zadaniem ponieważ metryka musi odpowiadać geometrii danych. Metryką skutecznie operującą na złożonych danych wielowymiarowych reprezentujących rozkład empiryczny cechy jest odległość Wassersteina. Jest ona niezwykle zbieżna z intuicyjnym rozumieniem braku podobieństwa rozkładów empirycznych. Określa bowiem coś co można nazwać minimalnym kosztem przekształcenia jednego rozkładu w drugi [123], [124]. Okazuje się, że daje to możliwość skutecznego analizowania zbiorów obrazów [125].

Niech (\mathcal{M}, d) będzie przestrzenią metryczną. Załóżmy, że każda miara prawdopodobieństwa zdefiniowana na tej przestrzeni jest regularna i skończona na odpowiednich podzbiorach \mathcal{M} . Niech $x_0 \in \mathcal{M}$ i

$$P(\mathcal{M}) = \{\mu : \mathcal{M} \rightarrow [0, 1] : \int_{\mathcal{M}} d(x, x_0) d\mu(x) < \infty\}. \quad (3.8)$$

Dla $\mu, \nu \in P(\mathcal{M})$ definiujemy *odległość Wassersteina* jako

$$W(\mu, \nu) := \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\mathcal{M} \times \mathcal{M}} d(x, y) d\gamma(x, y), \quad (3.9)$$

gdzie $\Gamma(\mu, \nu)$ jest zbiorem wszystkich miar na $\mathcal{M} \times \mathcal{M}$ takich, że μ i ν są odpowiednio ich rozkładami brzegowymi (marginałami). Jako funkcję kosztu d w niniejszej pracy przyjęto odległość Euklidesową.

Leonid Wasersztajn (Леонид Васерштейн) to rosyjski matematyk na cześć którego metryka została nazwana. W literaturze odwołującej się do niej najczęściej używana jest

zangielszczona wersja jego nazwiska - metryka Wassersteina. Wykorzystał on ją w swojej pracy z roku 1969 [126], jednak jako pierwszy zdefiniował ją Leonid Kantorowicz (Леонид Канторóвич) w 1939 roku [127]. Metryka Wassersteina znajduje wiele praktycznych zastosowań w analizie topologii danych, co odzwierciedlają liczne publikacje naukowe z różnych dziedzin. Np. praca [128] omawia temat uczenia metryk opisując ten proces jako trenowanie metryk. Celem algorytmu uczenia metryki jest skalibrowanie metryki, tak aby przypisywała małą odległość podobnym obiektom i stosunkowo dużą odległość istotnie różnym obiektom. Autorzy proponują wykorzystanie odległości Mahalanobisa jako funkcji kosztu do obliczeń metryki Wassersteina i automatyczne uczenie się odpowiednich macierzy za pomocą alternatywnego podejścia iteracyjnego. Autor przedstawia dokładność klasyfikacji różnych metod działających na danych wideo zawierających zapis ruchu drogowego i pokazuje, że jego metody przewyższają inne, takie jak metoda K-najbliższych sąsiadów wykorzystująca odległość euklidesową [129]. Publikacja [130] przedstawia wydajne algorytmy obliczania barycentrów Wassersteina jednocześnie pokazując, że jest to dobrze przystosowane narzędzie do generalizacji zbiorów obrazów 2D o różnorodnej treści. Dodatkowo autor [131] skonstruował dwupróbkowy test oparty na metryce Wassersteina, który jest przeznaczony do wykrywania strukturalnych wyłomów w danych o złożonej geometrii. Publikacja [132] opisuje metodologię śledzenia wielu pojazdów opartą na metryce asocjacyjnej Wassersteina. Opisane wyniki wskazują, że odległość Wassersteina dobrze sprawdza się w zadaniach wymagających oceny podobieństwa obiektów tego samego typu i jest odporna na częściową okluzję. Praca [133] przedstawia możliwość naturalnej generalizacji metryki Wassersteina na przypadek porównywania więcej niż dwóch rozkładów. Autorzy [134] opisują, jak wykorzystać generatywne sieci współzawodniczące Wassersteina (WGAN) do odsumiania i poprawy jakości obrazowania pozytonowej tomografii emisyjnej (PET) zmniejszając narażenie pacjenta na promieniowanie. Odległość Wassersteina zastosowana jako funkcja straty rozwiązuje problem nadmiernej gładkości i utraty szczegółowych informacji dotyczących klasycznych metod rekonstrukcji obrazu. Praca [135] udowadnia, że metryka Wassersteina może być wykorzystana jako dyskryminator jakości zrekonstruowanego obrazowania rezonansu magnetycznego. Wszystkie te badania dostarczają argumentów, że miara ta sprawdza się w zadaniach związanych z badaniem natężenia różnego rodzaju zjawisk. Dane o trajektoriach generowane przez uczestników ruchu są rozumiane jako systematyczna okupacja fragmentu przestrzeni względem autonomicznego pojazdu i stanowią kluczową treść scenariuszy testowych. Taka interpretacja nasuwa pomysł użycia narzędzia wrażliwego na ciągłe zmiany natężenia i lokalizacji danych. Dlatego w niniejszej pracy stawia się tezę, że odległość Wassersteina może stanowić podstawę do

opracowania metodologii, która pozwala na analizę dużych ilości danych, w szczególności składających się z trajektorii. Dokładniej, powinna ona umożliwić porównanie generowanego przez nie obciążenia na rozparcelowanej przestrzeni wokół samochodu (na siatce zajętości) pomiędzy różnymi scenariuszami. Oznacza to traktowanie zbiorów trajektorii jako pojedynczych obiektów bez utraty kluczowych informacji przez nieodpowiednią generalizację i w konsekwencji powinno umożliwić porównanie i charakteryzację zbiorów scenariuszy.

Rozdział 4

Nowa miara podobieństwa prostokątnych regionów rozpoznania

Aby skonstruować efektywną miarę podobieństwa dla porównania prostokątnych regionów rozpoznania, proponuje się w tej pracy trzy miary składowe, które mogą oceniać podobieństwo prostokątów pod względem atrybutów najistotniejszych dla oceny jakości rozpoznania. Później w oparciu o te funkcje konstruuje się ogólną miarę podobieństwa, która spełni pierwszy 3.2 oraz trzeci 3.4 warunek wymieniony w sekcji 3.2.1. Warunek dotyczący symetryczności 3.3 będzie spełniony wyłącznie przy odpowiedniej parametryzacji miary opartej na dystansie. Zaprojektowane miary składowe powinny być jak najbardziej niezależne od siebie. Postawiono taki cel, ponieważ każdą różnicę między dwoma prostokątami można w pełni opisać jako kombinację kilku prostych przekształceń geometrycznych, a zatem mając odpowiednio niezależne miary tych odchyłeń można przekształcić jeden prostokąt na drugi.

4.1. Podobieństwo prostokątnych obszarów rozpoznania

Niech \mathcal{R} będzie zbiorem wszystkich prostokątów w \mathbb{R}^2 . Zaprezentowane poniżej podejście kładzie istotny nacisk na fakt, że spośród dwóch porównywanych prostokątów jeden jest prostokątem wzorcowym (referencją) natomiast drugi jest kandydatem zaproponowanym przez testowany algorytm percepcji. Dlatego miary zaproponowane poniżej, a w szczególności miara podobieństwa pozycji oparta o odległość pomiędzy środkiem prostokąta wzorcowego i środkiem ocenianego prostokąta są zdefiniowane na przestrzeni

krzyżowej zbioru $\mathcal{R} \times \{0, 1\}$ z nim samym. Ustala się konwencję, że w każdej parze $(r, 0)$ prostokąt $r \in \mathcal{R}$ jest prostokątem wzorcowym stworzonym manualnie przez człowieka. W każdej parze $(r, 1)$ prostokąt r pochodzi z modułu percepcji. W oparciu o taką definicję przestrzeni na jakiej rozpatrywane w niniejszej pracy miary podobieństwa są położone uznaje się również, że dla dowolnej $f : (\mathcal{R} \times \{0, 1\}) \times (\mathcal{R} \times \{0, 1\}) \rightarrow \mathbb{R}$

$$f((r, n), (r', n')) = \begin{cases} f((r, n), (r', n')) & \text{gdy } n = 0 \wedge n' = 1 \\ f((r', n'), (r, n)) & \text{gdy } n = 1 \wedge n' = 0 \\ \frac{1}{2} & \text{gdy } n = n' \end{cases} . \quad (4.1)$$

Przyjmuje się takie oznaczenie by nie wywołać sprzeczności z warunkiem symetryczności 3.3, który w takim ujęciu problemu nie jest istotny. Aby skrócić zapis w kolejnych sekcjach uznaje się, że analizowane są wyłącznie wartości funkcji podobieństwa w przypadku gdy $n = 0 \wedge n' = 1$. Oznacza się L jako prostokąt referencyjny pochodzący z \mathcal{R} natomiast R jako prostokąt z \mathcal{R} będący kandydatem wybranym przez moduł percepcji.

4.1.1. Podobieństwo pola

Dla dwóch prostokątów $L, R \in \mathcal{R}$, definiuje się ich *podobieństwo pola* jako:

$$A(L, R) = \frac{\min(P(L), P(R))}{\max(P(L), P(R))}, \quad (4.2)$$

gdzie $P : \mathcal{R} \rightarrow [0, \infty)$ oznacza wartość pola danego prostokąta. To oznacza, że warunkiem koniecznym i wystarczającym aby prostokąty były pod tym względem w 100% podobne jest aby miały dokładnie takie samo pole. Ich kształt i pozycja nie mają wpływu na wartości tej funkcji.

4.1.2. Podobieństwo kształtu

W tej pracy rozpatrywane są prostokąty o tej własności, że każdy z nich ma dwa boki równoległe do pewnej ustalonej prostej. W przypadku detekcji obiektów na obrazie z kamery wynika to z tego, że każdy region detekcji jest prostokątem o bokach równoległych do krawędzi zdjęcia. W przypadku danych radarowych mamy do czynienia z widokiem z lotu ptaka i porównywane prostokąty mogą być ułożone względem siebie pod pewnymi kątami. Zawsze można jednak jeden bok wyróżnić stwierdzając, że reprezentuje np. przód obiektu (przedni zderzak w przypadku samochodów). Umożliwia to ustalenie konwencji dotyczącej kąta między przekątną, a bokiem prostokąta będącym jego szerokością. Uznaje

się, że ten kąt definiuje kształt prostokąta. Dla dwóch prostokątów $L, R \in \mathcal{R}$, definiuje się ich *podobieństwo kształtu* jako

$$S(L, R) = \cos^p(\alpha - \beta), \quad (4.3)$$

gdzie α i β to kąty definiujące kształty odpowiednio prostokąta L i R . Parametr $p \in \mathbb{R}^+$ służy redukcji wartości funkcji w zależności od potrzeby uwypuklania drobnych odchyleń w kształcie. Miara jest równa 100% jedynie w sytuacji, gdy wspomniane przekątne będą równoległe. Jakikolwiek odchylenie od tego stanu spowoduje obniżenie wartości podobieństwa. Pozycja i rozmiar prostokątów nie mają wpływu na wartości tej funkcji.

4.1.3. Podobieństwo pozycji

Niech (\mathbb{R}^2, d) będzie przestrzenią metryczną. Funkcję $s(\cdot)$ która prowadzi z \mathbb{R} w przedział $[0, 1]$ możemy nazywać *miarą podobieństwa opartą o metrykę d* jeśli dla dowolnych $x, y, z \in \mathbb{R}^2$,

$$d(x, y) \geq d(x, z) \Rightarrow s(d(x, y)) \leq s(d(x, z)). \quad (4.4)$$

Dodatkową własnością, która byłaby pożądana dla miary podobieństwa opartej na dystansie jest

$$d(x, y) = 0 \Leftrightarrow s(d(x, y)) = 1. \quad (4.5)$$

Niech $m : \mathcal{R} \rightarrow \mathbb{R}^2$ będzie funkcją, która zwraca punkt będący środkiem przekątnej dla danego prostokąta. Sprowadzenie badanych prostokątów do ich środka ciężkości uniezależnia miarę od kształtu i rozmiarów porównywanych prostokątów. Ta formalna niezależność zostanie później celowo zredukowana do lokalnej niezależności na pewnych podzbiorach $\mathcal{R} \times \mathcal{R}$ w celu uzyskania wartości, które będą odwoływać się do intuicyjnego rozumienia odległości między obiektami. Dla dwóch prostokątów $L, R \in \mathcal{R}$, można zdefiniować ich *podobieństwo pozycji* jako wartość funkcji

$$D(L, R) = \exp(-\gamma d(m(L), m(R))^\delta), \quad (4.6)$$

gdzie d jest metryką Euklidesową. Łatwo sprawdzić, że dla dodatnich i stałych parametrów γ i δ , powyżej wspomniane podstawowe warunki (4.4) i (4.5) są spełnione. Jednakże, funkcja podobieństwa pozycji, zdefiniowana przez równanie (4.6), bierze pod uwagę odległość Euklidesową środków porównywanych prostokątów. Tymczasem odległości realnych obiektów, takich jak np. piesi czy samochody, które reprezentowane są przez prostokąty,

skalowane są zgodnie z zasadami perspektywy. Podobieństwa pola i kształtu są wzajemnie niezależne czyli zmiana wartości jednej z tych miar nie powoduje zmiany wartości drugiej. Innymi słowy niezależność miary oznacza, że można podawać przykłady prostokątów o dowolnej wartości podobieństwa jednej z miar i jednocześnie mających stałą wartość drugiej i na odwrót. Ta sama właściwość charakteryzuje podobieństwo pozycji przy stałych parametrach γ i δ . W praktyce istnieje potrzeba uwzględnienia wpływu perspektywy na odległości pikseli zdjęcia. Obiekty znajdujące się w dużej odległości od kamery powinny być oceniane analogicznie do obiektów blisko kamery, pomimo znacznie mniejszej odległości Euklidesowej między pikselami wyznaczającymi ich środek ciężkości. Dlatego Autor niniejszej pracy zaproponował skalowanie miary odległości między centrami na podstawie rozmiarów porównywanych obiektów. Oznacza to pośrednią zależność miary dystansu od rozmiaru porównywanych prostokątów. Metoda skalowania jest opisana poniżej, opiera się ona na praktycznych potrzebach i możliwości uniwersalnego interpretowania zmienności KPI.

Parametry γ i δ skalują wartości podobieństwa pozycji dla dopasowań poprawnych, krytycznie złych oraz zapewniają płynne i dobrze opisane przejście przez przypadki pośrednie. Aby zdefiniować γ, δ , zakłada się, że oczekujemy od funkcji podobieństwa aby osiągała wartość s_1 gdy dystans pomiędzy środkami porównywanych prostokątów $d(L, R)$ jest równy p_1 i wartość s_2 gdy ten dystans jest równy p_2 . Podstawowe założenia o tych parametrach są następujące

$$p_1, p_2 : \mathcal{R} \times \mathcal{R} \rightarrow \mathbb{R}_0^+ \quad \wedge \quad \forall_{L, R \in \mathcal{R}} p_1(L, R) > p_2(L, R), \quad (4.7a)$$

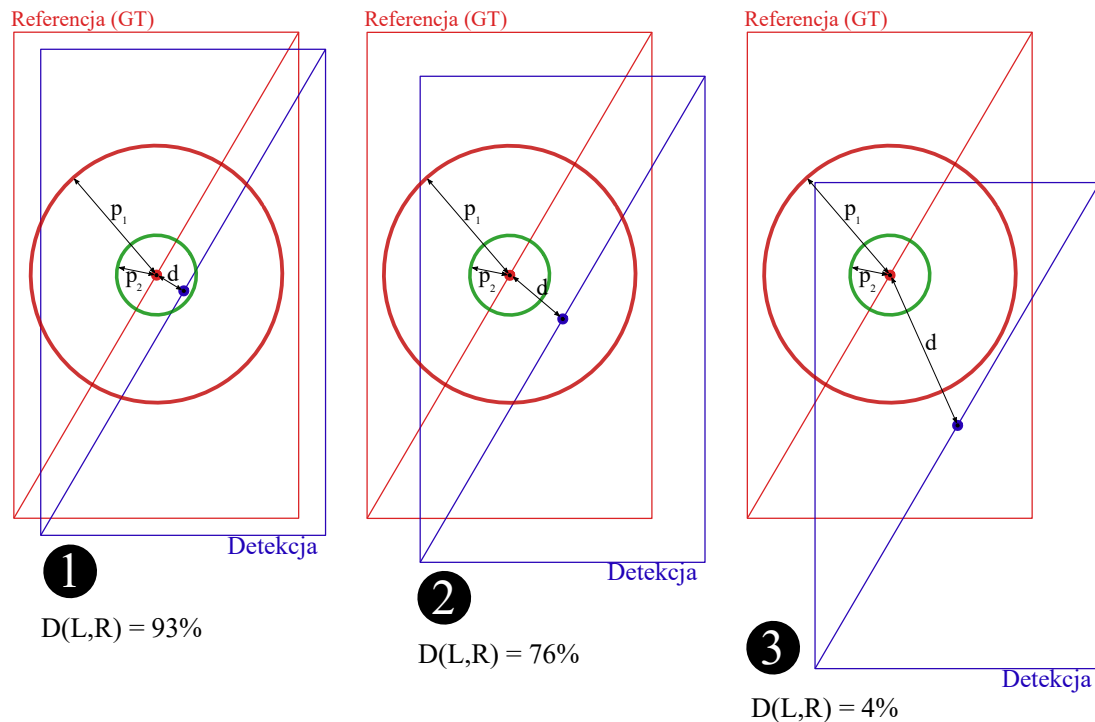
$$s_1, s_2 \in (0, 1) \quad \wedge \quad 0 < s_1 < s_2. \quad (4.7b)$$

Wyznaczenie wzorów na γ i δ wymaga rozwiązania następującego układu równań

$$\begin{cases} s_1 = e^{-\gamma p_1(L, R)^\delta}, \\ s_2 = e^{-\gamma p_2(L, R)^\delta}, \end{cases} \quad (4.8)$$

dzięki czemu otrzymuje się

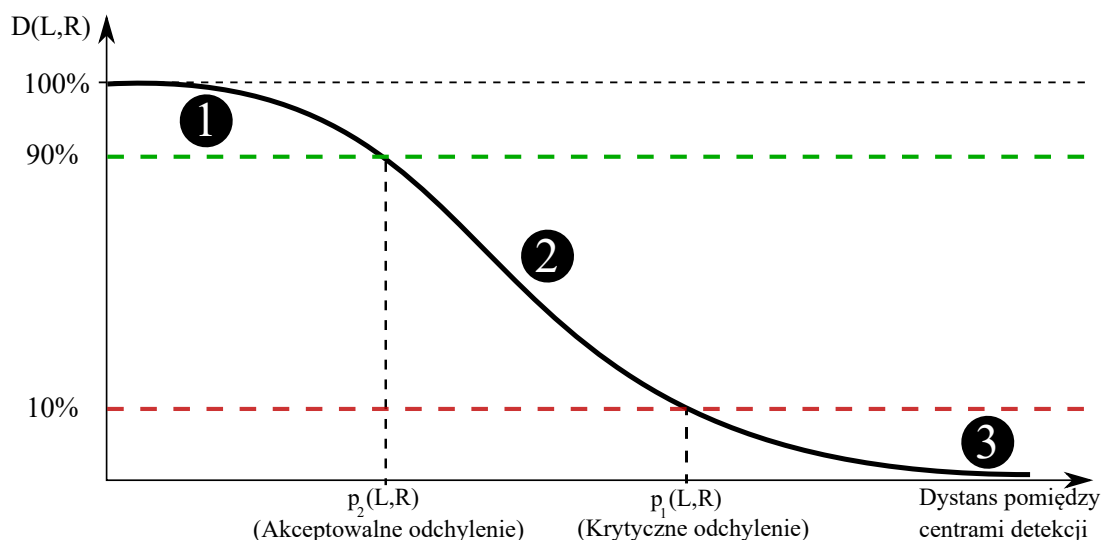
$$\begin{cases} \delta(L, R) = \ln \frac{\ln s_1 / \ln \frac{p_1(L, R)}{p_2(L, R)}}{-\ln s_1}, \\ \gamma(L, R) = \frac{1}{p_1(L, R)^{\delta(L, R)}}. \end{cases} \quad (4.9)$$



Rysunek 4.1. Podobieństwo pozycji reprezentuje różnicę położenia prostokątów odniesioną do rozmiarów porównywanych obiektów

Powyższe równania pozwalają wyliczyć współczynniki δ i γ aby jednolicie skalować wartości miary podobieństwa wobec wartości odległości między centrami detekcji prostokątów (rysunek 4.1). Skalowanie bazujące na wartościach s_1, s_2, p_1, p_2 , pozwala jasno zdefiniować jakie dopasowania rozumiane są jako akceptowalne oraz krytycznie złe, co w praktyce ułatwia komunikowanie wyników bez potrzeby ponownego wyjaśniania opisanych tu szczegółów technicznych. Finałowym krokiem jest skalibrowanie miary podobieństwa pozycji czyli wyznaczenie zależności między parametrami p_1, p_2 , a rozmiarami porównywanych prostokątów. Funkcje $p_1(L, R)$ i $p_2(L, R)$ powinny uwzględniać specyfikę danego modułu percepcji. Dodatkowo, z punktu widzenia produkcyjnego, kalibracja powinna uwzględniać wymagania klienta. Proponowane przykłady kalibracji dla kluczowych modułów percepcji są podane w kolejnych sekcjach.

Warto zwrócić uwagę na to, że symetryczność miary podobieństwa pozycji byłaby uwarunkowana symetrycznością funkcji $p_1(L, R)$ i $p_2(L, R)$ gdyby nie założenia poczynione w 4.1. Motywacją do rezygnacji z symetrycznej kalibracji jest żądanie, aby prostokąty będące referencją miały znacznie istotniejszy wpływ na dobór krzywizny funkcji oceny podobieństwa pozycji (rysunek 4.2). Jak się okaże dalej, dzięki temu możliwa jest



Rysunek 4.2. Wynik pomiaru jakości pozycji uzyskuje się po przeskalowaniu krzywej funkcji indywidualnie dla każdego porównania. Parametry p_1 i p_2 opisują główne progi oceny właściwości prostokątnych obszarów rozpoznania

realna kontrola nad niezależnym wpływem wyników referencyjnych i wyników pochodzących z czujników na zaproponowaną metodę oceny dobrych i złych dopasowań. Wszystkie pośrednie obserwacje są oceniane na podstawie jednej uniwersalnej krzywej, danej równaniem (4.6), która skaluje się indywidualnie dla każdego porównania. Biorąc to pod uwagę należy zauważyć, że aby zapewnić wypukłość tej krzywej w celu utrzymania wysokich wartości miary w zakresie małych wartości odległości, parametr δ powinien spełniać dodatkowe założenie $\delta > 1$. W praktyce dopóki s_1 i s_2 są od siebie znacznie oddalone, co jest naturalne jeżeli wartości „poprawne” i „niepoprawne” mają się bardzo różnić, inżynier ma dużą swobodę w kwestii wyboru parametrów p_1 , p_2 , nawet biorąc pod uwagę, że muszą spełniać warunki (4.7a) oraz (4.7b).

4.1.4. Generalna miara podobieństwa

Bazując na trzech zdefiniowanych składowych miarach podobieństwa, można zaproponować generalną miarę podobieństwa prostokątów, scalającą informacje niesione przez dotychczas przedstawione funkcje podobieństwa. Miara ta spełnia podstawowe warunki definicji miary podobieństwa 3.2 i 3.3 oraz posiada własność dodatkową 3.4. Dla $L, R \in \mathcal{R}$ definiuje się generalną miarę podobieństwa prostokątów GMOS (ang. *general measure of similarity*) jako

$$G(L, R) = \frac{3}{\frac{w_1}{S(L, R)} + \frac{w_2}{A(L, R)} + \frac{w_3}{D(L, R)}}, \quad (4.10)$$

Miara ta przyjmuje wartość 100% wtedy i tylko wtedy gdy prostokąty są identyczne w sensie wszystkich trzech miar składowych jednocześnie. Ważona średnia harmoniczna została wybrana, ponieważ jest najbardziej wrażliwa na wartości odstające spośród trzech najpopularniejszych średnich: arytmetycznej, geometrycznej i harmonicznej. Jednocześnie przy dominującej roli miary poprawnej pozycji pozostałe miary mają pełnić rolę pomocniczą. Aby osiągnąć podobny efekt np. wykorzystując średnią arytmetyczną, wagi dla miar kształtu i pola należałoby wybrać skrajnie niskie w porównaniu do wagi dla miary podobieństwa pozycji. W efekcie stałyby się zupełnie nieistotne, a to nie jest celem Autora. Zaletą takiego podejścia jest możliwość dokładnego zdefiniowania tego, co można uznać za prawidłowe rozpoznanie zmniejszając istotnie liczbę przypadków wątpliwych w procesie automatycznego oceniania.

4.1.5. Dopasowanie detekcji między źródłami

W pojedynczej klatce filmu może znajdować się wiele obiektów, których wykrycie jest przewidziane w danych referencyjnych. Jednocześnie zbiór prostokątnych regionów rozpoznania, jaki moduł percepcji zaproponuje dla danej klatki, może zawierać inną liczbę elementów. Zanim dokonamy pomiaru jakości wykrycia wszystkich obiektów musimy zdecydować, czy dla każdego prostokąta referencyjnego istnieje prostokąt pochodzący z systemu i przyporządkować go tak, aby nie został wykorzystany do pomiaru jakości rozpoznania innego obiektu w klatce. Aby skutecznie przeprowadzić takie przyporządkowanie, warto wykorzystać informacje, jakie dostarczają poszczególne składowe generalnej miary podobieństwa. Algorytm dopasowania (asocjacji) składa się z dwóch kroków. Dla każdego obiektu referencyjnego L_i ze zbioru $\{L_1, L_2, \dots, L_n\}$ znajdujemy podzbiór kandydatów spośród dostępnych ze zbioru prostokątów dostarczonych przez testowany moduł percepcji R_1, R_2, \dots, R_k . Do kandydatów dla L_i dodajemy wszystkie detekcje R_j nie przyporządkowane wcześniej innemu obiektowi referencyjnemu oraz spełniające warunki minimalne:

$$\begin{cases} A(L_i, R_j) \geq 0,25, \\ S(L_i, R_j) \geq 0,9, \\ G(L_i, R_j) \geq 0,1. \end{cases} \quad (4.11)$$

Następnie spośród tak wybranych kandydatów wybieramy jednego o najwyższej wartości miary D pod warunkiem, że nie ma w klatce innego obiektu referencyjnego, dla którego dany kandydat spełnia warunki minimalne i jest bardziej podobny w sensie miary D . W przypadku równości decyduje miara A .

Algorytm 1 Asocjacja prostokątów między źródłami

```

1: for  $i = 1, 2, \dots, n$  do
2:   for  $j = 1, 2, \dots, k$  do
3:     if  $L_i$  i  $R_j$  spełniają warunki minimalne 4.11  $\wedge R_j$  nie został jeszcze przyporządkowany then
4:        $R_j$  dołącza do zbioru kandydatów  $C_i$ 
5:     end if
6:   end for
7:   for  $R \in C_i$  do
8:     if  $\exists k \in \{1, 2, \dots, n\} : L_k$  i  $R$  spełniają warunki minimalne 4.11  $\wedge D(L_k, R) > D(L_i, R)$  then
9:        $C_i = C_i - \{R\}$ 
10:    end if
11:  end for
12:  if  $C_i = \emptyset$  then
13:     $L_i$  jest sklasyfikowany jako rozpoznanie FP
14:  else if  $|C_i| = 1$  then
15:    Jedyne kandydat z  $C_i$  zostaje przyporządkowany do  $L_i$ 
16:  else
17:    if Istnieje dokładnie jeden kandydat  $R' \in C_i$  realizujący  $\max\{D(L_i, R) : R \in C_i\}$  then
18:       $R'$  zostaje przyporządkowany do  $L_i$ 
19:    else
20:      W podzbiorze kandydatów realizujących  $\max\{D(L_i, R) : R \in C_i\}$  do  $L_i$  zostaje przyporządkowany ten posiadający najwyższą wartość miary  $A$ 
21:    end if
22:  end if
23: end for

```

Podane warunki minimalne są szczególnie użyteczne w przypadku dużej gęstości detekcji jak na przykład w przypadku zatłoczonego przejścia dla pieszych, które zawsze stanowi wyzwanie dla modułu PED. Dla modułów, dla których liczba detekcji w klatce jest mocno ograniczona (MOD) lub odległości między obiektami są znaczne w porównaniu do ich rozmiarów (TSR) można zrezygnować z warunków minimalnych i oprzeć algorytm dopasowujący detekcje do obiektów tylko na wartościach miary D .

Aby przeanalizować dokładność algorytmu przeprowadzono eksperyment na ponad 100

klatkach przedstawiających przechodniów. Dla tych klatek przygotowano dane referencyjne oraz przykładowy zestaw detekcji. W sumie w rzeczonych klatkach analizowanych detekcji było 713. Na takim zbiorze danych uruchomiono zaproponowany algorytm asocjacji i manualnie przeanalizowano wyniki znajdując liczbę poprawnie zidentyfikowanych rozpoznań prawdziwie pozytywnych. Dla porównania, równolegle, zastosowano też algorytm dopasowujący detekcje i GT na podstawie wartości indeksu Jaccarda. Prawdziwie pozytywny wynik był raportowany w tym przypadku dla wartości $J(L, R) > 30\%$. Indeks Jaccarda uzyskał w tym eksperymencie wynik 94.3% poprawnie zaklasyfikowanych detekcji prawdziwie pozytywnych. Zaproponowany powyżej algorytm bazujący na miarach podobieństwa prostokątów uzyskał dokładność na poziomie 98.2%.

Obniżenie progu dla indeksu Jaccarda powoduje ryzyko zaakceptowania krytycznie złego rozpoznania jako prawdziwie pozytywnego co jest szczegółowo omówione w sekcji 4.4. Może to prowadzić do przeszacowania precyzji detektora. Statystyka ta jest niewrażliwa na informacje o wynikach jakości lokalnej, dlatego w dalszej części zaproponowano alternatywny sposób podsumowania jakości lokalnej dla sekwencji detekcji, która korzysta bezpośrednio z wartości pomiaru lokalnej jakości rozpoznań prawdziwie pozytywnych.

4.2. Analiza zdarzeń w czasie

Zaproponowane miary podobieństwa pozwalają na skuteczne wyszukanie ze zbioru prostokątów wskazanych przez badany system percepcji tego, który można powiązać z zadanym obiektem referencyjnym. W praktyce oznacza to, że w każdej ramce filmu każdy prostokąt referencyjny może być z dużą dozą pewności prawidłowo oznaczony jako wykryty lub niewykryty, a każda prawdziwie pozytywna detekcja ma przypisaną generalną miarę jakości. Zakłada się, że obiekty niewykryte (nie mające dopasowanego odpowiednika w zbiorze prostokątów zaproponowanych przez badany system percepcji) mają jakość wykrycia równą 0% dla wszystkich miar. Nadmiarowe prostokąty pochodzące z modułu percepcji oznaczają się jako wykrycia fałszywie pozytywne. Każdy referencyjny prostokątny region rozpoznania w pojedynczej ramce filmu jest najczęściej kolejnym w całej sekwencji prostokątów, które w kolejnych klatkach opisują pojedynczy obiekt pojawiający się w zasięgu czujnika pojazdu autonomicznego. Innymi słowy możemy rozumieć, że pojedynczy prostokąt referencyjny jest dwuwymiarowym przekrojem trójwymiarowej bryły reprezentującej obiekt poruszający się przez pewien dyskretnie próbkowany w polu widzenia kamery (lub innego czujnika). Oznacza to, że można z zestawu kolejno występujących po sobie klatek filmu wydzielić spójną sekwencję prostokątów, opisujących zdarzenie polega-

jące na wykryciu i rejestracji obiektu, który pojawił się w pobliżu samochodu. Kolejnym zadaniem ewaluacyjnym jest określenie - dla każdego wykrytego obiektu - miary jakości wykrycia obiektu w całym czasie jego obecności w zasięgu czujników. Miarę tę ustala się zwykle na podstawie wszystkich lokalnych wartości miary jakości rozpoznania obiektu w kolejnych klatkach filmu [136]. Jest to kolejny krok określania jakości działania modułu percepcji. Aby ocenić jakość wykrycia pojedynczego obiektu w czasie jego przebywania w zasięgu sensorów pojazdu, wyposażonego w moduły percepcji, należy w jakiś formalny sposób podsumować wartości miar podobieństwa w zakresie ramek, dla których istnieje prostokąt referencyjny opisujący ten obiekt.

4.2.1. Miara podobieństwa sekwencji

Niech sekwencja prostokątów referencyjnych pojawiających się w serii następujących po sobie klatek filmu $\mathcal{L} = \{L_1, \dots, L_n\}$ reprezentuje obiekt, który znajdował się w zasięgu czujników dla kolejnych n klatek. Przekrój obiektu znajdujący się w klatce o indeksie i jest prostokątem referencyjnym $L_i \in \mathcal{R}$. Niech $I \subset \{1, \dots, n\}$ będzie zbiorem tych indeksów klatek, w których istnieje prostokąt pochodzący z modułu percepcji połączony z odpowiednim prostokątem L_i , $i \in I$, (czyli zbiór i gdzie istnieje $R_i \in \mathcal{R}$, który ma obrazować ten sam obiekt co L_i). Niech $\mathcal{P} = \{R_k, \dots, R_l\}$ będzie zbiorem tych wyników działania badanego modułu percepcji, a pochodzących z różnych klatek, które są połączone z odpowiednim L_i dla $i \in I$. Definiuje się miarę podobieństwa sekwencji \mathcal{G} , która porównuje ciąg referencyjnych prostokątów \mathcal{L} z ciągiem przyporządkowanych im prostokątów wygenerowanych przez oceniany moduł percepcji \mathcal{P} w następujący sposób:

$$\mathcal{G}(\mathcal{L}, \mathcal{P}) = \frac{1}{|\mathcal{L}|} \sum_{i=1}^{|\mathcal{L}|} w_i o(i), \quad (4.12a)$$

gdzie

$$o(i) = \begin{cases} G(L_i, R_i) & \text{dla } i \in I \\ 0 & \text{w pozostałych przypadkach} \end{cases}. \quad (4.12b)$$

Wartości w_i są wagami wyznaczonymi na bazie algorytmu uwzględniającego dwa, istotne z praktycznego punktu widzenia, aspekty:

1. Minimalny czas jaki może być wymagany by ewaluowany moduł percepcji zareagował na nowo pojawiający się obiekt.

2. Kluczowa dla bezpieczeństwa lokalizacja momentu pierwszego poprawnego rozpoznania względem pierwszego pojawienia się obiektu w zasięgu czujników.

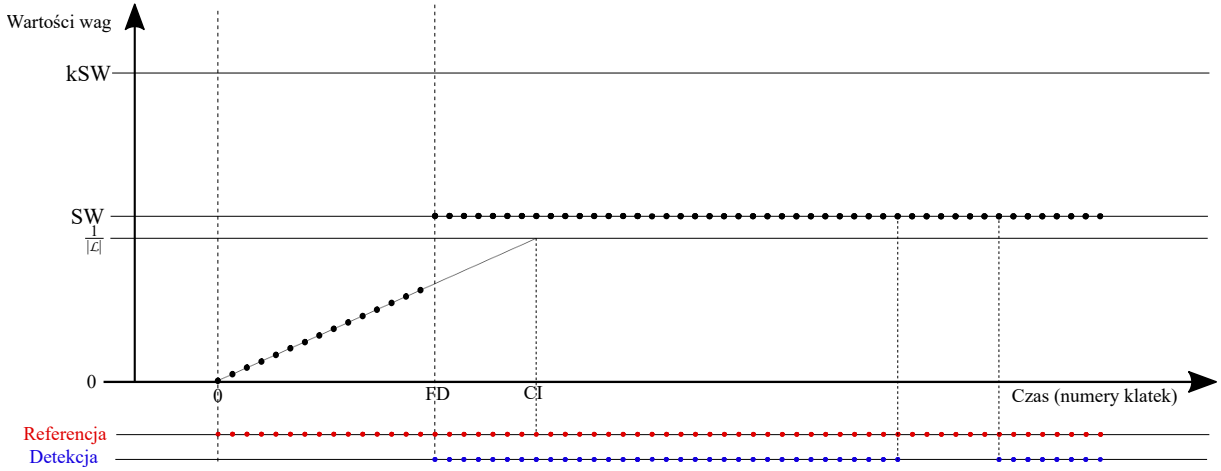
4.2.2. Średnia ważona dla późnych detekcji

Stosunkowo często się zdarza, że obiekt musi pozostawać w zasięgu czujników przez pewien czas zanim algorytm dokona pierwszej udanej detekcji. Pierwszy prostokąt dla danego obiektu jest pierwszym wykryciem. Jest to typowe zachowanie modułów detekcji, dlatego metoda ewaluacji musi je uwzględnić. Aby ograniczyć negatywny wpływ opóźnienia detekcji na średnią arytmetyczną podsumowującą jakość detekcji, użyto wag. Wydaje się, że z oczywistych względów kluczowym dla bezpieczeństwa ruchu drogowego jest jak najwcześniejsze wykrycie nowego obiektu, który znajduje się w zasięgu czujników pojazdu. Dlatego metoda ewaluacji powinna dodatkowo uwzględniać zwłokę wykraczającą poza założoną i tolerowaną wartość opóźnienia. Wagi zostaną zatem wybrane, tak aby uwypuklić nadmierną wartość opóźnienia detekcji obiektów i aby upewnić się, że w tym przypadku miara podobieństwa sekwencji obniży się istotnie. Te dwa warunki skłaniają do zdefiniowania dwuczęściowego algorytmu obliczania wag. W tym celu wprowadzmy oznaczenia:

- *Indeks krytyczny (CI)* to liczba klatek, licząc od momentu pojawienia się obiektu w zasięgu czujników, w czasie których nie oczekuje się detekcji obiektu.
- *Pierwsza detekcja (FD)* to numer klatki, licząc od momentu pojawienia się obiektu w zasięgu czujników, w którym pierwszy raz do prostokąta referencyjnego opisującego obiekt, został dopasowany prostokąt pochodzący z modułu detekcji i można wyliczyć wartość generalnej miary jakości detekcji na podstawie miar podobieństwa prostokątów.
- *Waga standardowa (SW)* to waga przyporządkowana każdej wartości miary jakości detekcji, obliczanej od momentu pierwszej detekcji do końca sekwencji \mathcal{L} .

Metoda obliczenia zestawu wag dla danej sekwencji jest uzależniona od czasowej lokalizacji pierwszej detekcji względem indeksu krytycznego. Ze względu na to, rozpatrujemy dwa przypadki wzoru dobierania wag:

1. Pierwszy przypadek zachodzi, gdy pierwsza detekcja znajduje się przed indeksem krytycznym lub jest mu równa – rysunek 4.3. Dla indeksów od 1 do CI , tworzy się funkcję liniową, która wartość 0 przyjmuje dla 1 i wartość $|\mathcal{L}|^{-1}$ dla indeksu



Rysunek 4.3. Pierwszy sposób obliczania wag dla kolejnych indeksów sekwencji gdy $FD \leq CI$

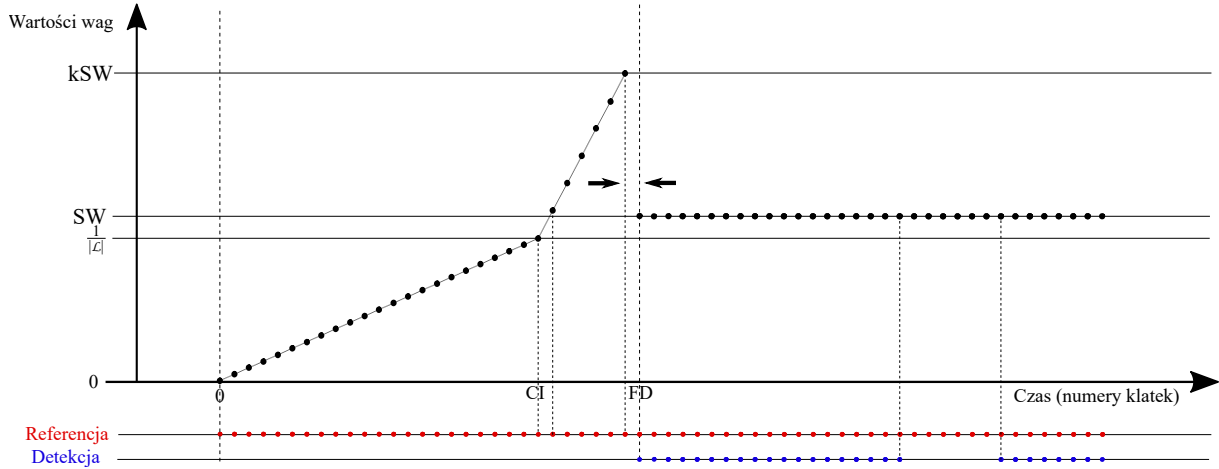
CI . Wagi dla indeksów od 1 do $FD - 1$ są równe wartości funkcji dla tych indeksów. Waga z indeksem FD oraz wszystkie pozostałe wagi są równe SW , nawet jeśli algorytm zgubi detekcję (czyli jeśli jakiś podciąg późniejszych referencyjnych prostokątów nie będzie miał dopasowanych prostokątów z algorytmu wizyjnego). Oznacza to, że późniejszy brak detekcji ma zbalansowany wpływ analogiczny do pozostałych obserwacji. Niewielkie wartości wagi dobierane na początku przed indeksem krytycznym sprawiają, że brak detekcji na samym początku pojawienia się obiektu nie ma dużego wpływu na obniżenie się wartości miary podobieństwa sekwencji.

$$w_i = \begin{cases} i \cdot \frac{|\mathcal{L}|^{-1}}{CI} & \text{dla } 0 \leq i < FD \\ SW & \text{w pozostałych przypadkach} \end{cases}, \quad (4.13a)$$

gdzie

$$SW = \frac{|\mathcal{L}| - \sum_{i=0}^{FD-1} w_i}{|\mathcal{L}| - FD + 1}. \quad (4.13b)$$

2. Drugi przypadek zachodzi gdy pierwsza detekcja pojawia się po indeksie krytycznym – rysunek 4.4. W tym przypadku dobiera się wagi dla indeksów od 1 do CI , korzystając z podejścia opisanego w przypadku pierwszym, czyli waga z indeksem



Rysunek 4.4. Drugi sposób obliczania wag dla kolejnych indeksów sekwencji gdy $FD > CI$

CI jest równa $|\mathcal{L}|^{-1}$. Suma wag w tym przedziale równa się zatem

$$I_1 = \sum_{i=1}^{CI} i \cdot \frac{|\mathcal{L}|^{-1}}{CI}. \quad (4.14)$$

Wagi z kolejnymi indeksami to wartości dla tych indeksów funkcji liniowej, która dla argumentu CI przyjmuje wartość $|\mathcal{L}|^{-1}$ i dla argumentu $FD - 1$ przyjmuje wartość $k \cdot SW$. Parametr $k > 1$ dobiera się w zależności od tego, jak mocno chcemy karać spóźniającą się pierwszą detekcją. Suma wag w tym przedziale jest równa

$$I_2 = \sum_{i=CI+1}^{FD-1} \left(\frac{k \cdot SW - |\mathcal{L}|^{-1}}{FD - CI} \cdot i + |\mathcal{L}|^{-1} - \frac{k \cdot SW - |\mathcal{L}|^{-1}}{FD - CI} \cdot CI \right). \quad (4.15)$$

Wagi z indeksem FD i wszystkie pozostałe są równe SW , którą to wartość należy obliczyć z równania liniowego

$$|\mathcal{L}| = I_1 + I_2 + (|\mathcal{L}| - FD + 1) \cdot SW, \quad (4.16)$$

ponieważ każdy zestaw wag musi sumować się do $|\mathcal{L}|$ (liczba klatek w jakich pojawiła się referencja związana z danym obiektem). Gwarantuje to, że tworzona funkcja będzie miarą podobieństwa oraz ocena jakości, dla dowolnie długich par sekwencji \mathcal{L} , \mathcal{P} , będzie mogła być bezpośrednio porównywana z wynikiem dla krótkich sekwencji.

Ostatecznie standardową wagę można wprost wyliczyć ze wzoru

$$SW = \frac{|\mathcal{L}|^2 - \sum_{i=1}^{CI} \frac{1}{CI} + \sum_{i=CI+1}^{FD-1} \frac{i}{FD - CI} + \left(\frac{2CI - FD}{FD - CI}\right)(FD - CI - 2)}{|\mathcal{L}| \cdot k \cdot \sum_{i=CI+1}^{FD-1} \frac{i}{FD - CI} + \frac{|\mathcal{L}| \cdot k}{FD - CI}(FD - CI - 2) + |\mathcal{L}|(|\mathcal{L}| - FD + 1)} \quad (4.17)$$

W sytuacji gdy SW jest niedodatnie:

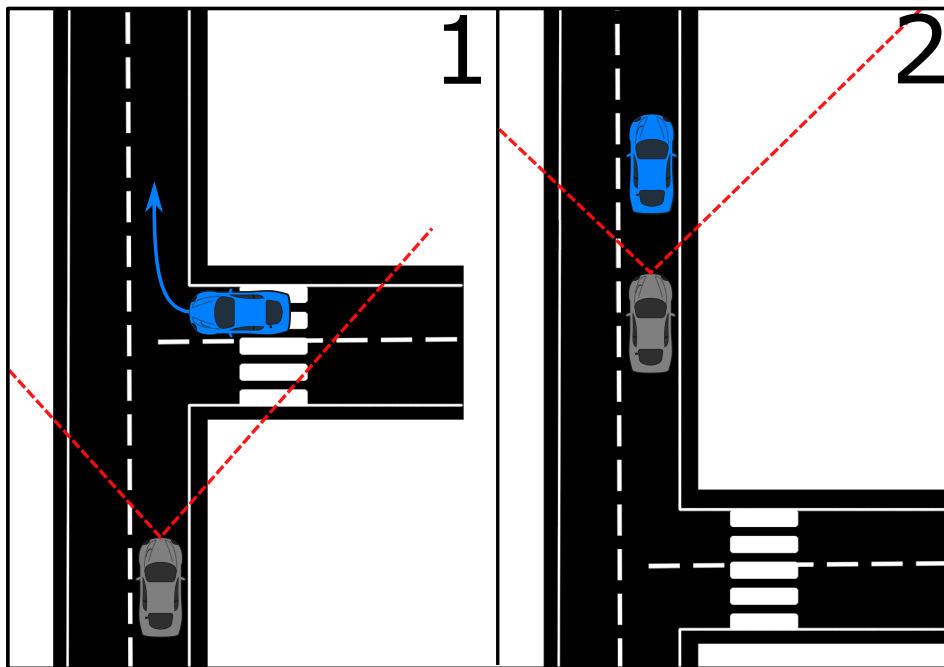
$$\mathcal{G}(\mathcal{L}, \mathcal{P}) = 0. \quad (4.18)$$

Można też dynamicznie manipulować parametrem CI w zależności od sytuacji na drodze. Przykładowo, można uzależnić CI od prędkości samochodu posiadającego inteligentną percepcję w danym momencie i uwzględnić potrzebę szybszego wykrycia obiektów przy większych prędkościach. Można też uwzględnić utrudnienia powodowane przez warunki pogodowe, takie jak np. mgła i zwiększyć tolerancję dla późniejszych wykryć. Manipulowanie parametrem CI daje ciekawe możliwości. W niniejszej pracy skupiono się jednak na testowaniu miary podobieństwa pozycji dla stałych wartości parametrów $CI = 24$ i $k = 2$.

4.2.3. Rola pierwszej detekcji

Standardowo dane wizyjne w samochodach testowych zapisywane są jako filmy pochodzące z kamer o prędkości zapisu 30 FPS, co oznacza, że liczba pojedynczych klatek wideo, które kamera przechwytuje na sekundę jest równa 30. Część z czasu akwizycji ramki wideo należy przyjąć jako wkład w opóźnienie algorytmów wizyjnych zaimplementowanych w samochodzie. Ten efekt w opisanym algorytmie uwzględniony jest przez parametr CI .

Przyjęta w tej pracy wartość parametru CI odpowiada opóźnieniu detekcji krótszemu niż 1 sekunda. Można zapytać: dlaczego istnieje potrzeba by surowo karać oceniany moduł, obniżając wartość miary podobieństwa, w przypadku gdy pierwsze wykrycie przekracza CI ? Motywacją takiego podejścia z punktu widzenia bezpieczeństwa ruchu drogowego jest fakt, że pojawienie się nowego obiektu w otoczeniu samochodu bardzo często odpowiada znaczącej zmianie sytuacji na drodze. Wiąże się to z potrzebą podejmowania szybkich decyzji przez protokoły kontrolne w samochodzie i zapewnienia, że interpreta-

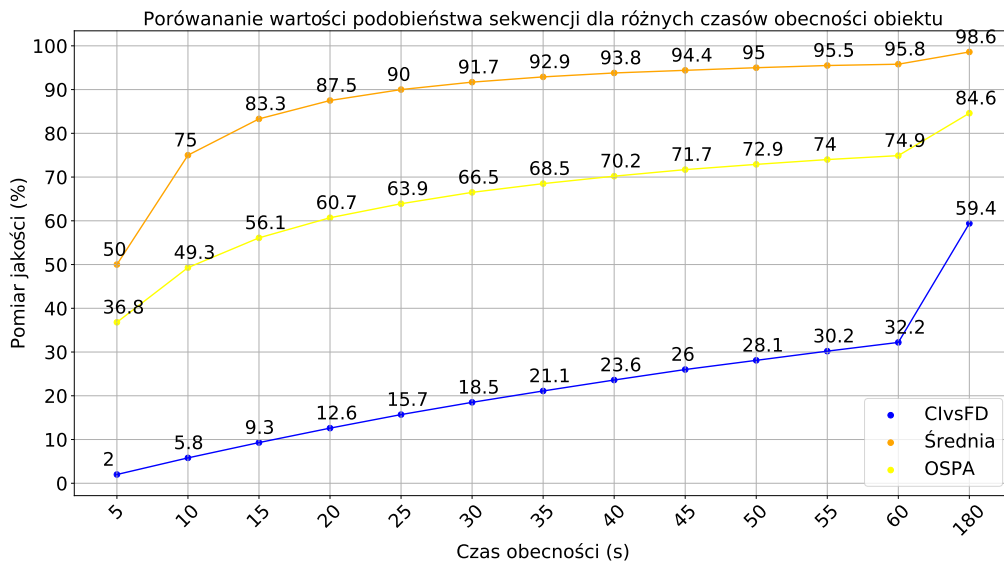


Rysunek 4.5. Przykład sytuacji na drodze, kiedy spóźnione pierwsze wykrycie może odpowiadać pominięciu kluczowego zdarzenia podczas całego pojawienia się obiektu; 1 – włączenie się nowego obiektu do ruchu; 2 – podążanie za obiektem. Nowe podejście do oceny śledzenia zostało skonstruowane, tak aby uwrażliwić KPI na tę sytuację i odpowiednio podkreślić jej istotność podczas podsumowania ewaluacji

cja otoczenia pojazdu odpowiada, w tym momencie, możliwie dokładnie rzeczywistości. Więcej informacji na ten temat można znaleźć w [137] i [138]. Fakt, że podczas jazdy samochodem wydłużony o ułamki sekundy czas reakcji może spowodować drastyczne zmiany w efektach wykonywanego manewru, nie powinien wymagać dodatkowych wyjaśnień [139]. Jazdy testowe są prowadzone w taki sposób, że nawet w sytuacji, gdy algorytm sterowania zawiedzie, to człowiek ma szansę zapewnić skuteczną egzekucję poprawnych manewrów, a dane są nadal gromadzone. Nagranie z jazdy testowej jest później wielokrotnie wykorzystane wtórnie, by symulować działanie nowszych wersji algorytmów wykrywania w celu pomiaru ich jakości.

Aby unaocznić problem motywujący do takiego podejścia, przeanalizujmy hipotetyczną sytuację. Zakładając, że używa się klasycznej średniej dla podsumowania wszystkich pomiarów jakości z wideo, rozważono sytuację przedstawioną na rysunku 4.5. Na lewej części rysunku widoczny jest samochód, który na skrzyżowaniu wjeżdża w pole widzenia pojazdu autonomicznego. Przyjęto, że algorytm przeoczy go podczas większej części lub całego opisanego manewru, który najprawdopodobniej wymagałby natychmia-

stowego dostosowania prędkości lub zmiany pasa przez samochód z tyłu. Podczas symulacji w tym momencie przez kilka sekund otrzymywane będą zera z miar podobieństwa. Następnie sytuacja na drodze stabilizuje się, samochód testowy podąża za obiektem, który pojawił się przed nim. W tym momencie przyjęto, że algorytm wizyjny generuje pierwszą prawidłową detekcję i raportuje ją od tego punktu w czasie. Obiekt porusza się przed samochodem testowym przez kilka minut prawidłowo rozpoznany, jak widać na drugiej połowie rysunku 4.5. Klasyczna średnia wartość wszelkiego rodzaju miar podobieństwa, zebranych dla tego obiektu, będzie wysoka ze względu na dysproporcję prawidłowego rozpoznania z drugiej fazy sceny względem pierwszej, co zmniejsza wpływ zer z początkowego kontaktu. Taki rezultat jest zdecydowanie niesatysfakcjonujący. Proponowany w tej pracy sposób ważenia średniej rozwiązuje problem. Aby numerycznie zilustrować skuteczność proponowanej metody, obliczono miarę jakości opisanej wyżej serii detekcji. Przyjęto, że w lokalnym ujęciu miara jakości rozpoznania przyjmie jedynie wartości 0% lub 100%, po to aby skupić się wyłącznie na wpływie na ocenę doboru miary podobieństwa sekwencji. Porównano analizę wydajności śledzenia z wykorzystaniem standardowego podejścia do średniej, metryki OSPA i opisanej w tej sekcji autorskiej metody (nazwanej *CIvsFP*). Założono, że przedstawiony po lewej stronie rysunku 4.5 manewr włączenia się do ruchu trwa 5 sekund. Oznacza to, że otrzyma się 150 wyników pomiaru jakości dla wszystkich klatek w tym okresie. Ponadto przyjęto, że moduł percepcji nie zlokalizował obiektu przez początkowe 2,5 sekundy co skutkuje otrzymaniem 75 klatek o wartości lokalnej miary dla tego obiektu równej 0%. Po czym moduł zadziałał idealnie do samego końca obecności obiektu, co skutkuje uzyskaniem 75 klatek z miarą 100%. W tym momencie dokonano pierwszego pomiaru jakości przy założeniu, że to koniec sekwencji. Następnie stopniowo przedłuża się fazę drugą sztucznej sytuacji (prawa strona rysunku.4.5). Dodano po 5 sekund w każdym kroku, aż osiągnięto całkowity czas detekcji równy jedną minutę. W każdym kroku zmierzono jakość całego zdarzenia wszystkimi metodami (przy tymczasowym założeniu $CI = 0,1s$. aby skupić się na efekcie *FD*). Dodatkowo wykonano dla tego zdarzenia analizę uzupełniającą, trwającą 3 minuty (ponieważ jest to maksymalna długość standardowego scenariusza testowego). Wyniki przedstawiono na rysunku 4.6. Klasyczna wartość średnia wstępnie podsumowuje jakość tego rozpoznania jako 50% i szybko wzrasta do wartości powyżej 90% wraz z wydłużeniem się fazy pozostawiania obiektu w zasięgu czujników, sugerując w ostatecznych wynikach, że mamy do czynienia z niemal idealnym wykryciem. Algorytm OSPA zachowuje się lepiej jednak niewystarczająco dobrze. W rzeczywistości nieudane wykrycie, które powoduje poważne zagrożenie, jest maskowane przez prawidłowe wykrycia z późniejszego okresu śledzenia obiektu. W zasadzie odwrotny efekt



Rysunek 4.6. Porównanie analizy jakości śledzenia obiektu przy pomocy zaproponowanej metody, standardowej średniej oraz miary OSPA

można zaobserwować, analizując jak wbudowana ochrona w zaproponowanej metodzie ClvsFD zapobiega uzyskiwaniu mylących wyników w takiej sytuacji. Ostatnia obserwacja pochodzi ze sceny, która trwa 3 minuty i jest granicą długości scenariusza. Warto podkreślić, że podejście to w naturalny sposób koncentruje się na zastosowaniach produkcyjnych w motoryzacji.

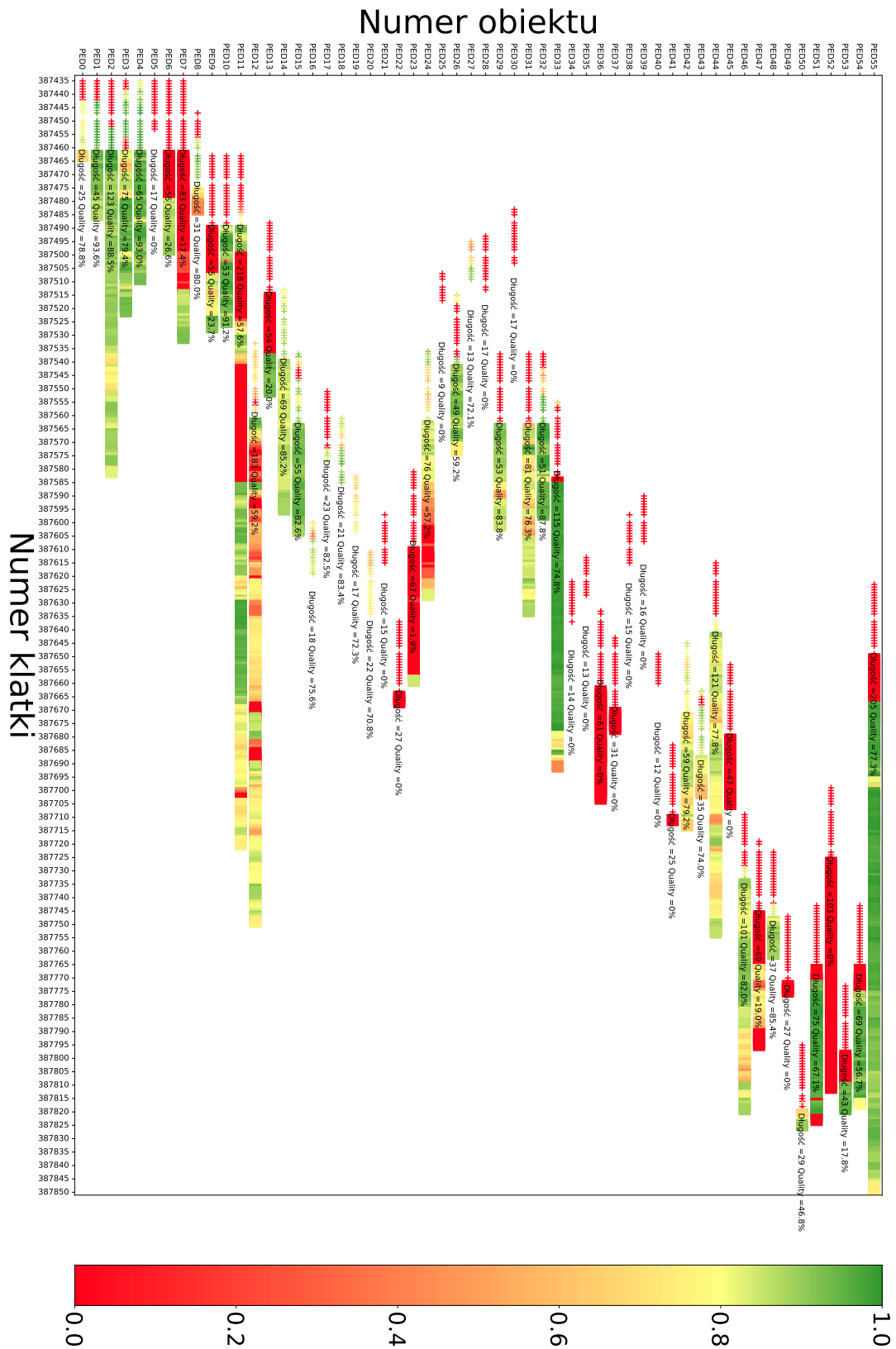
4.3. Wizualizacja jakości sceny testowej

Poniżej zaprezentowano formę wizualizacji wyników analizy jakości zaprojektowanej na potrzeby przykładowego modułu percepcji (PED). Przejrzysta i bogata wizualizacja jakości rozpoznań w scenie testowej ma istotne znaczenie nie tylko ze względu na możliwość szybkiego i precyzyjnego lokalizowania problemów modułów percepcji, ale również podnosi skuteczność komunikacji wewnątrz zespołu badawczego oraz na linii dostawca - klient. Na wykresie 4.7 pozioma oś reprezentuje czas mierzony w numerach klatek. Pionowa oś zawiera identyfikator obiektu, dlatego każdy poziomy pas reprezentuje jakość pojedynczego obiektu, podczas jego pojawienia się w polu widzenia kamery. Znaczniki krzyża oznaczają stan detekcji przed indeksem krytycznym dla danego zdarzenia. Kolory pojedynczych stanów są dobierane na podstawie wartości generalnej miary podobieństwa

GMOS natomiast na pasku podane są wartości $\mathcal{G}(\mathcal{L}, \mathcal{P})$ jakie algorytm przypisał dla danego obiektu. Taka reprezentacja danych pozwala na szybką lokalizację problematycznych sytuacji na filmie, dostarcza informacji, jakie obiekty pojawiły się na drodze jednocześnie i potencjalnie weszły w interakcję oraz daje możliwość prześledzić jak jakość wykrycia zmieniała się w czasie.

4.3.1. Rozpoznania fałszywie pozytywne

Wyniki fałszywie pozytywne to te prostokątne regiony rozpoznania, pochodzące z modułów percepcji, które nie zostały w procesie łączenia dopasowane do żadnych prostokątów referencyjnych w kontekście pojedynczych ramek. Jeśli fałszywe rozpoznania pojawiają się w danym obszarze tylko przez krótki czas, można uznać je za nieszkodliwe i zignorować podczas analizy jakości modułu. Samochód otrzymujący ciąglą fałszywą informację o obecności obiektu w pobliżu, może podjąć nieprzewidywalne i najprawdopodobniej błędne decyzje. Podjęte przy złych założeniach działanie ma dużą szansę na doprowadzenie do niebezpiecznego zdarzenia na drodze. Dlatego przedłużające się w czasie fałszywie pozytywne ciągi rozpoznań powinno się traktować ponownie jako spójne zdarzenie i uznawać za potencjalnie szkodliwe. Użyteczną do tego celu będzie zaproponowana wcześniej miara podobieństwa pozycji. Pozwoli ona dopasować wyniki fałszywie pozytywne między kolejnymi klatkami. Na potrzeby tego zadania proponuje się porzucić symetryczność uzyskiwaną przez założenia z 4.1 i dobrać symetryczne parametry p_1 i p_2 . Zwykle konieczne jest też zmniejszenie wagi podobieństwa pozycji we wzorze 4.10, aby połączyć w jeden ciąg niestabilne detekcje. Wreszcie, jeśli nie można znaleźć podobnych wyników fałszywie dodatnich w następnej klatce (lub kilku klatkach), zamyka się zdarzenie i zgłasza jego długość, średnią szerokość, wysokość i pozycję. Dodatkowo można sprawdzić, czy początek tego zdarzenia fałszywie pozytywnego ma swoje korzenie w GT, które zakończyło się w ostatniej klatce, przed pojawieniem się początkowego rozpoznania sklasyfikowanego jako fałszywie pozytywne (podobnie dla wyników zakorzenionych na początku zdarzenia). Zazwyczaj takie sytuacje również nie muszą być rozpatrywane jako niebezpieczne. Przykładowa wizualizacja wyników fałszywie pozytywnych znajduje się rysunku 4.8. Podobnie jak w przypadku poprzedniej wizualizacji, poziome pasy dotyczą osobnych obiektów fałszywie pozytywnych. Kolor szary służy do oznaczenia nieszkodliwych, fałszywych wyników pojawiających się krótko. W kolorze czerwonym to potencjalnie istotne zaburzenia, pojawiające się długo, które nie są związane z referencją. Zakorzenione w re-



Rysunek 4.7. Wizualizacja jakości rozpoznania kolejno pojawiających się w zasięgu czujników obiektów oceniona przy pomocy GMOS i miary podobieństwa sekwencji

ferencji obiekty fałszywie pozytywne oznaczone są kolorem pomarańczowym. Wielkość i kształt znaczników informuje o szerokości prostokątów i ich lokalizacji na klatce filmu (lewa strona, centrum i prawa strona). Dzięki takiemu podejściu można przeprowadzić automatyczną analizę wstępną i pomóc w ustaleniu priorytetów, co znacznie przyspieszy z kolei analizę manualną lub automatyczną, jeśli będzie potrzebna w kolejnym kroku.

4.4. Kalibracja miar jakości dopasowania

Poniżej przedstawiono przykładowe wyniki analizy jakości rozpoznania w klatce dla najistotniejszych modułów percepcji. Przytacza się również specyficzną kalibrację wybraną w celu przystosowania zaprojektowanych miar jakości do charakterystycznych wyzwań, jakie stoją na drodze do poprawnej automatycznej ewaluacji tych modułów.

4.4.1. Wymagania kalibracyjne wybranych modułów

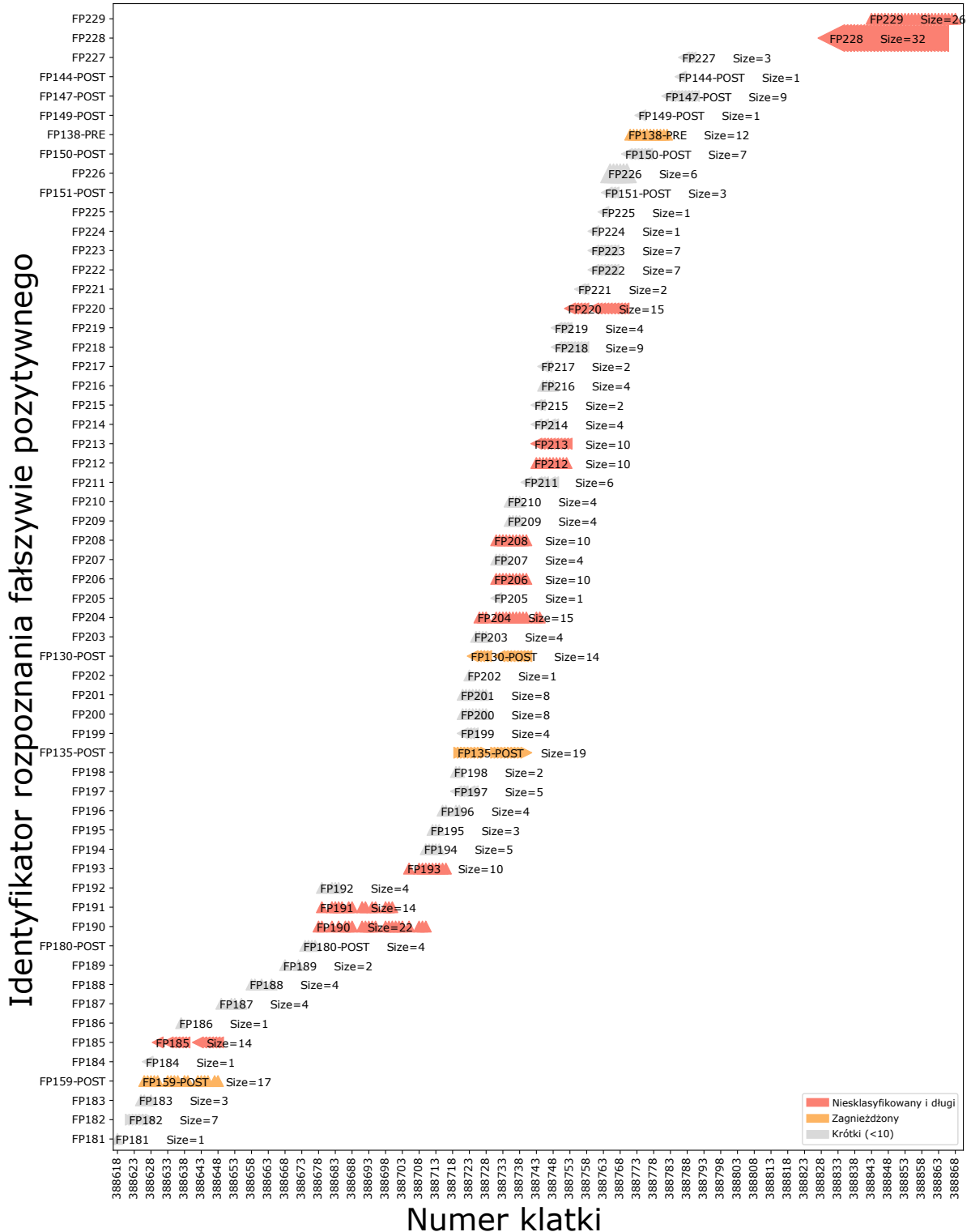
PED

Do analizy tych danych wykorzystano miary zdefiniowane w 4.1, przyjmując we wzorach 4.9 następujące parametry kalibracyjne:

$$\begin{aligned} s_1 &= 10\%, s_2 = 90\%, \\ p_1(L, R) &= \frac{4}{10}\text{diag}(L) + \frac{2}{10}\text{diag}(R), \\ p_2(L, R) &= \frac{2}{10}\text{diag}(L) + \frac{1}{10}\text{diag}(R), \end{aligned}$$

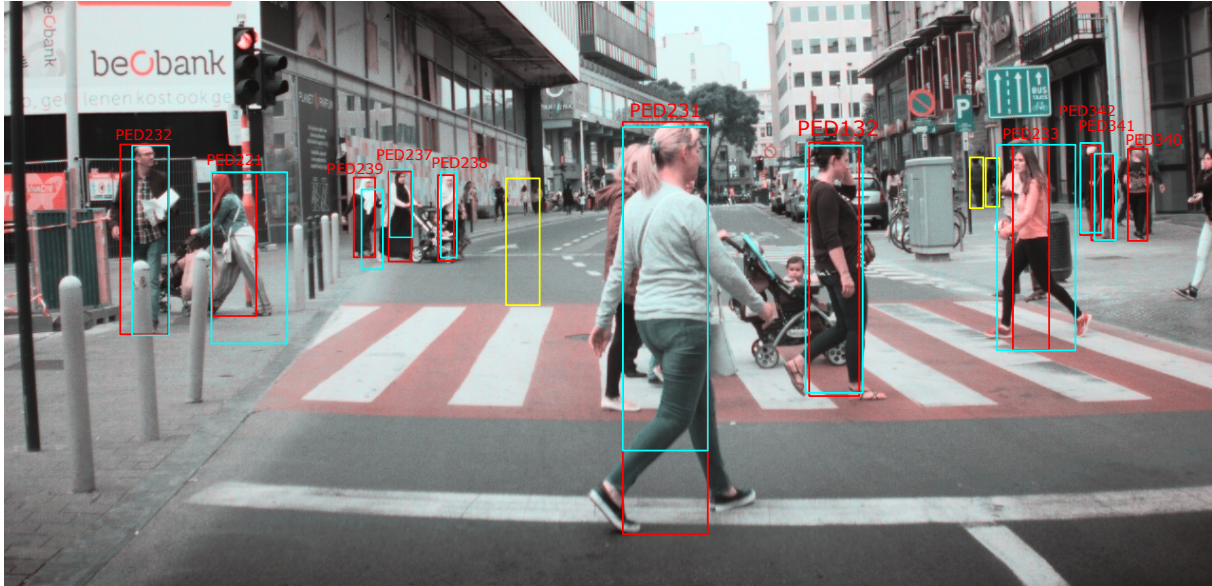
gdzie $\text{diag} : \mathcal{R} \rightarrow (0, \infty)$ to funkcja przyporządkowująca prostokątowi długość jego przekątnej. W ten sposób uzależniono skalowanie funkcji podobieństwa pozycji od rozmiarów prostokątów, kładąc nacisk na właściwości prostokątów referencyjnych. Wartości tak dobranych miar oraz wartości Indeksu Jaccarda (3.5), dla przykładów zawartych w klatce na rysunku 4.9, są przedstawione w tabeli 4.1.

W prezentowanym przykładzie prostokąt opisujący pieszego o numerze 221 ma złą pozycję, kształt i powierzchnię. Zarówno miara generalna, jak i indeks Jaccarda zwracają podobnie niską ocenę dla tego rozpoznania, przy czym niska wartość indeksu Jaccarda jest spowodowana wyłącznie zbyt dużym polem powierzchni prostokąta rozpoznania. Sytuacja pieszego z numerem 233 jest podobna, ale położenie prostokąta jest prawidłowe. Indeks Jaccarda ocenia to rozpoznanie jako prawie identyczne pod względem jakości, jak



Rysunek 4.8. Wizualizacja lokacji, klasyfikacji i opisu znalezionych obiektów fałszywie pozytywnych

poprzednio. Miara generalna nadal uznaje to rozpoznanie za złe, ale wartość oceny jest znacznie wyższa, ponieważ rzeczywiście ten wynik jest lepszy. Pieszego o numerze 239 opisuje prostokąt o odpowiednim polu powierzchni i kształcie, ale jego pozycja jest nieodpowiednia. Przesunięcie w stosunku do prostokąta referencyjnego (czerwony) należy uznać za znaczące, ze względu na dużą odległość obiektów od pojazdu mimo, że wyniki różnią się zaledwie kilkoma pikselami. Największym problemem stosowania indeksu Jaccarda w celu ewaluacji detekcji pieszych jest to, że wpływ na jakość błędnie wyznaczonej szerokości prostokąta jest mnożony przez wysokość obiektu. Samo ignorowanie wysokości nie jest niezawodnym rozwiązaniem, ponieważ może prowadzić do utraty informacji o pozycji. Podobieństwo pola boryka się z tym samym problemem (w sytuacji, gdy jeden prostokąt jest w pełni zawarty w drugim, wartości tej miary są dokładnie takie same jak wartości indeksu Jaccarda – patrz pieszy numer 233). Jednak możliwość osadzenia informacji w kontekście pozostałych miar składowych pozwala odróżnić sytuację, w której rozbieżność wartości pola detekcji i referencji jest rzeczywiście szkodliwa. Szczególną uwagę należy zwrócić na przypadek kluczowy z punktu protokołów sterowania pojazdów samochodowych – piesi o numerach 231 i 132. Oba przechodzą przez ulicę przed samochodem. Pieszy o numerze 231 znajduje się na linii wyznaczającej miejsce zatrzymania pojazdu przed przejściem, a 132 znajduje się na przejściu dla pieszych. Szerokość pieszego 231 jest doskonale opisana przez niebieski region rozpoznania będący detekcją, ale wysokość jest przez niego krytycznie źle opisana (szczególnie w dolnej partii), co może prowadzić do mylącego i niebezpiecznego wniosku, że osoba ta prawidłowo przechodzi przez ulicę. Detekcja przechodnia numer 132 jest obciążona niewielkim przesunięciem wszystkich krawędzi względem referencji ale nadal można go uznać za prawie doskonałe rozpoznanie. Index Jaccarda ocenia oba te przypadki niemal identycznie - 80,4% za wykrycie obciążone dużym błędem i 79,1% za prawie doskonałe wykrycie. Generalna miara podobieństwa prostokątów opisuje złą detekcję wartością 39,5%, a doskonałą detekcję wartością 94,5%. Podobne porównanie można przeprowadzić dla pieszych 237 i 238, którzy pojawiają się w tle. GMOS ponownie dostarcza wartości, które ułatwiają prawidłowy proces decyzyjny. Oznacza to, że wartości tej miary są w dużym stopniu niezależne od skali. Jest to jedna z najbardziej pożądanых właściwości miary jakości detekcji.



Rysunek 4.9. Prostokąty referencyjne (czerwone) i prostokąty pochodzące z modułu percepcji PED (niebieskie - prawdziwie pozytywne i żółte - fałszywie pozytywne)

Tabela 4.1. Podobieństwo prostokątów pochodzących z referencji i modułu percepcji PED na rysunku 4.9

PED	Jaccard	GMOS	Pozycja	Pole	Kształt	Luminescencja
221	43,6%	41,3%	37,0%	43,6%	85,3%	25,5%
233	39,0%	63,3%	99,0%	39,0%	64,4%	97,7%
239	36,7%	47,4%	34,1%	98,6%	99,9%	08,0%
231	80,4%	39,5%	28,3%	80,4%	97,8%	29,7%
132	79,1%	94,5%	99,8%	85,8%	97,7%	99,1%
237	70,6%	35,9%	25,2%	80,7%	97,0%	67,0%
238	56,1%	86,4%	78,4%	100 %	100 %	78,0%
232	69,4%	86,8%	99,7%	69,4%	95,9%	99,7%
342	80,4%	97,4%	96,2%	98,9%	98,8%	97,3%
341	70,4%	97,8%	96,3%	100 %	100 %	71,3%
340	0%	0%	0%	0%	0%	0%

MOD

Innym problemem, który pojawia się w trakcie oceny dopasowania detekcji prostokątnych, jest to, że czasami można uzyskać wyniki zasadniczo odmienne od referencji, ale mimo wszystko powinno się je uznać za poprawne, podobnie jak te, które idealnie się z nią pokrywają. Miara podobieństwa pozycji okazuje się bardzo pomocna w tej sytuacji, ponieważ można modyfikować definicję centrum detekcji w zależności od regionu, dla którego określenie położenia jest szczególnie ważne (lub jeśli istnieje region, którego oznaczenie jest nieistotne). Dobrym przykładem takiej sytuacji jest wykrywanie wysokich pojazdów, takich jak ciężarówki. Czasami w nocy kamery samochodu wykrywają tylko dolną część ciężarówki, tą w której znajdują się tylne światła, zamiast wykrywać całą ciężarówkę (patrz rysunek 4.10). Ponieważ to wystarczy, aby układ sterujący samochodem zachowywał się prawidłowo, potrzeba miary, która pozytywnie oceni taką detekcję pomimo, że dopasowanie nie jest idealne. Aby uwzględnić takie sytuacje w algorytmie, do proponowanej oceny odległości dla tego modułu dodajemy możliwość obniżenia porównywanych centrów detekcji w stosunku do ich domyślnych pozycji. Najpierw należy rozpoznać, czy wysokość środka prostokąta detekcji znajduje się wyżej niż położenie środka odpowiedniego prostokąta referencyjnego. Jeśli to prawda, nie jest wymagana żadna korekta. W przeciwnym razie przesuwany jest środek detekcji obu prostokątów przed wyliczeniem wartości miary pozycji. Zmiana w danym prostokącie jest zależna od stosunku wysokości H do szerokości W prostokąta. Im wyższa wartość współczynnika, tym większe przesunięcie środka detekcji od środka geometrycznego prostokąta. Wartości przesunięcia mogą być oparte na krzywej przekształcenia odwrotnej funkcji logitowej, aby zachować pierwotną pozycję środka wykrywania obiektów, które nie są wysokie. Limit przesunięcia to jedna piąta wysokości prostokąta:

$$y_{\text{nowy}} = y_{\text{stary}} - \frac{H}{5 \cdot (1 + \exp(-H/W))}, \quad (4.19)$$

gdzie y_{stare} i y_{nowe} są odpowiednio wartością rzędnej środka detekcji przed i po korekcie. Te poprawki są stosowane niezależnie dla prostokąta referencyjnego i wyniku pochodzącego z modułu. Oznacza to, że gdy różnica wysokości między prostokątami jest niewielka, oba centra detekcji zostaną przesunięte w podobny sposób, a wartość funkcji podobieństwa nie zmieni się znacząco w stosunku do jej oryginalnej definicji. Gdy system rozpozna tylko dolną część referencji, przesunięcie środka detekcji nie będzie znaczące, natomiast środek referencji obniży się i znajdzie w jej dolnej części. Spowoduje to, że funkcja podobieństwa pozycji będzie miała rozsądną wartość. Informacja, że to dopasowanie nie jest idealne,

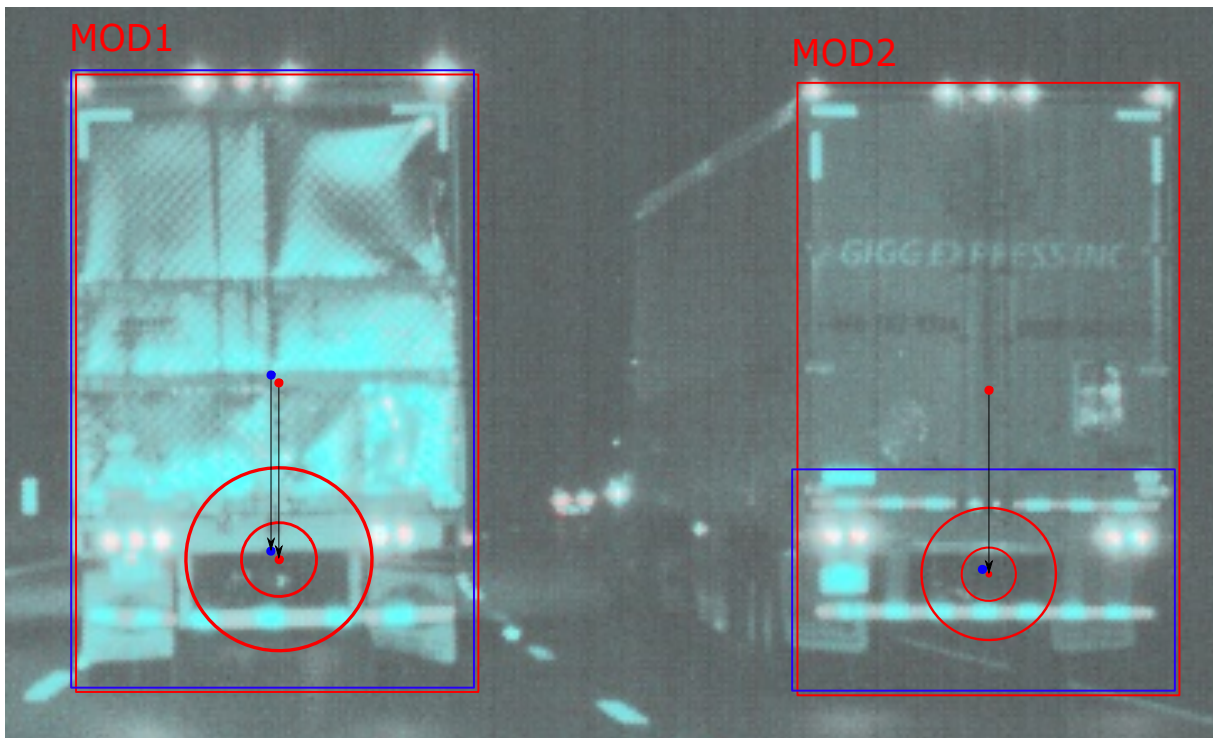
będzie widoczna po sprawdzeniu wartości podobieństwa powierzchni oraz wartości podobieństwa kształtu. Co się tyczy kalibrowania parametrów granicznych, to w przypadku tego modułu percepcji, z racji, że rozpoznaje się zwykle obiekty o dużych gabarytach, ograniczamy do minimum wpływ rozmiaru przekątnej detekcji na tolerancję przesunięcia samych środków względem siebie

$$s_1 = 10\%, s_2 = 90\%,$$

$$p_1(L, R) = \frac{3}{5}\text{diag}(L) + \frac{1}{16}\text{diag}(R),$$

$$p_2(L, R) = \frac{1}{18}\text{diag}(L).$$

Przykłady wartości miar o takiej kalibracji można zobaczyć w tabeli 4.2. Przesuwając środek każdego obiektu, koryguje się wartość podobieństwa odległości, a tym samym zapewnia, że wynik nie zostanie sklasyfikowany jako obiekt fałszywie pozytywny, a referencja nierozpoznana. Otrzymuje się informację o satysfakcjonująco rozpoznanym obiekcie.



Rysunek 4.10. Prostokąty referencyjne (czerwone) i prostokąty pochodzące z modułu percepcji MOD (niebieskie)

Tabela 4.2. Podobieństwo prostokątów pochodzących z referencji i modułu percepcji MOD na rysunku 4.10

MOD	Jaccard	General	Distance	Area	Shape	Luminescencja
1	94,4%	99,5%	99,4%	99,5%	100%	100%
2	32,7%	36,6%	98,5%	33,0%	9,1%	84,1%

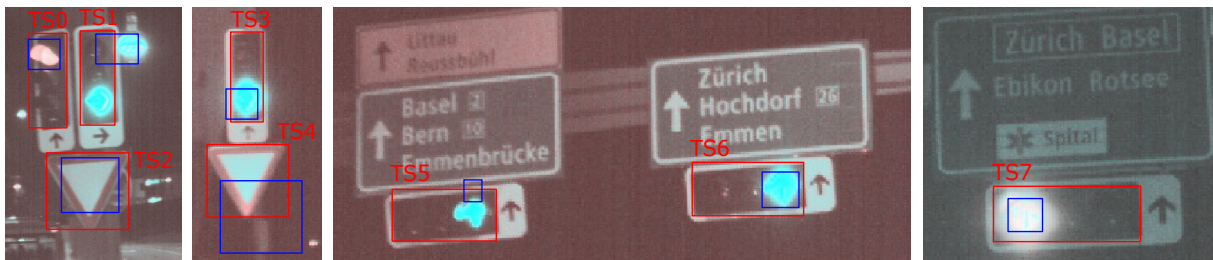
TSR

Podobny problem pojawia się w trakcie rozpoznawania sygnalizacji świetlnej. Referencja powstaje przy założeniu, że system percepcji oznaczy całą obudowę sygnalizatora, niezależnie od tego, jaki jest stan sygnalizacji. Jednak w nocy można się spodziewać, że czujniki rozpoznają tylko świecące fragmenty sygnalizatora. W takim przypadku istnieje potrzeba oceny wykrytego pola jako dopuszczalnego wyniku, o ile rozpoznawany jest iluminujący obszar. W ten sposób rozpoznawanie sygnalizacji świetlnej w dzień jest oceniane domyślnie, natomiast w nocy jest bardziej tolerancyjne. Aby to zrobić automatycznie, należy najpierw zdecydować, czy należy wyregulować centrum detekcji. Miara powierzchni i kształtu może dostarczyć tej informacji. Następnie, na podstawie struktury świateł (pionowej lub poziomej) i stanu luminescencji odnotowanej jako dodatkowy parametr referencji, algorytm może dostosować położenie centrum referencji i obliczyć podobieństwo odległości dla fragmentarycznej detekcji, aby poprawnie określić skuteczność wykrycia (porównaj rysunek 4.12 i tabela 4.3).

$$s_1 = 10\%, s_2 = 90\%,$$

$$p_1(L, R) = \frac{4}{5}\text{diag}(L),$$

$$p_2(L, R) = \frac{1}{5}\text{diag}(L).$$



Rysunek 4.11. Prostokąty referencyjne (czerwone) i prostokąty pochodzące z modułu percepcji TSR (niebieskie)

Tabela 4.3. Podobieństwo prostokątów pochodzących z referencji i modułu percepcji TSR na rysunku 4.12

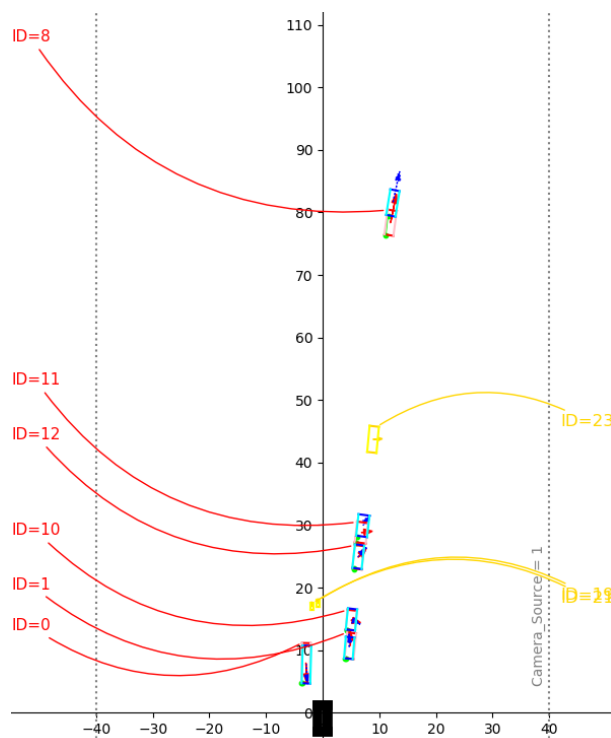
TS	Jaccard	GMOS	Pozycja	Pole	Kształt
0	21,7%	22,6%	99,8%	23,3%	8,6%
1	9,9%	4,5%	34,8%	28,2%	0,9%
2	41,7%	70,5%	96,3%	41,7%	100,0%
3	22,6%	40,7%	91,6%	31,6%	28,4%
4	26,5%	61,7%	46,6%	94,3%	99,3%
5	4,2%	13,4%	26,9%	10,5%	5,2%
6	19,0%	30,8%	100,0%	19,0%	25,3%
7	15,4%	26,0%	100,0%	15,4%	18,2%

4.4.2. Analiza danych radarowych i lidarowych

Przedstawioną w tym rozdziale metodologię można zaadaptować do analizy danych detekcji na podstawie wskazań czujników radarowych i lidarowych. Dla wyników radarowych dane pochodzące z lidarów mogą stanowić referencję, ponieważ są one uważane za znacznie dokładniejsze i bardziej odporne na niekorzystny wpływ otoczenia [2]. Metodologia ta może być również wykorzystana do sprawdzenia, czy wyniki detekcji dla systemu zainstalowanego w samochodzie testowym pokrywają się z danymi uzyskanymi na podstawie resymulacji, co jest ważne gdy zachodzi potrzeba dowodu reprodukowalności jakości uzyskiwanej w symulatorze w warunkach pracy w pojeździe.

Perspektywa lotu ptaka

Wyniki, które otrzymuje się z tych systemów, są ponownie w postaci prostokątów o określonej szerokości i długości, ale tym razem reprezentują obiekty wokół głównego pojazdu widziane z perspektywy lotu ptaka. Każdy wynik ma przypisany kierunek (odchylenie pojazdu od osi wyznaczonej przez kierunek samochodu testowego) oraz wektor prędkości, który reprezentuje kurs i względną prędkość obiektu. Do porównania położenia, wielkości i kształtu obiektów można wykorzystać zaproponowaną w tej pracy metodologię oceny prostokątów. Do oceny podobieństwa rotacji bryły pojazdu można wykorzystać funkcję analogiczną do wzoru (4.3), ale jako argument należy podać różnicę w kątach determinujących rotację w wynikach pochodzących z porównywanych źródeł. Taką miarę można zaimplementować jako kolejny składnik miary GMOS. Warto zwrócić uwagę, że w większości przypadków kąt odchylenia jest uzależniony od wektora prędkości.



Rysunek 4.12. Prostokąty lidarowe (czerwone) i prostokąty radarowe (niebieskie - prawdziwie pozytywne i żółte - fałszywie pozytywne). Strzałki reprezentują wektor prędkości względnej obiektu

Miara dopasowania prędkości

Aby ocenić podobieństwo między wektorami prędkości dla wyniku radarowego \vec{v}_r i wyniku lidarowego \vec{v}_l , można ponownie użyć funkcji analogicznej do miary podobieństwa pozycji, jednak interpretacja wartości jest zupełnie inna. Argumentem dla tej funkcji jest długość różnicy wektorów $|\vec{v}_l - \vec{v}_r|$. Parametry $p_1(\vec{v}_l, \vec{v}_r)$ i $p_2(\vec{v}_l, \vec{v}_r)$ można wybrać na podstawie długości wektorów (czyli na podstawie szybkości względnej obiektów, do których są one przypisane). Wtedy miarą podobieństwa prędkości będzie

$$V(\vec{v}_l, \vec{v}_r) = \exp\left(-\gamma(d(\vec{v}_l, \vec{v}_r))^\delta\right). \quad (4.20)$$

Gdy kierunek obu wektorów jest taki sam, argumentem tej funkcji jest surowa różnica między prędkościami obiektów. Taką samą wartość tej miary uzyskuje się, gdy różnica kursu pojawi się jednocześnie z odpowiednim zmniejszeniem różnicy prędkości skalarnych. Innymi słowy, funkcja ta przyjmuje tę samą wartość w sytuacjach, gdy na podstawie przypisanych wektorów prędkości obiekt hipotetycznie znajdzie się po jednostkowym odstępie czasu w pozycjach oddalonych o tę samą odległość. Dzięki takiej definicji otrzymujemy

miarę, która normalizuje błędy różnego rodzaju w pełnym kontekście prędkości pojazdów, umożliwiając jednolitą ocenę jakości.

4.5. Porównanie miar jakości lokalnej

W tej sekcji porównano zaproponowaną miarę podobieństwa prostokątów z innymi miarami dostępnymi w literaturze. Na początku porównano ogólne własności, jakie można przypisać wybranym miarom, a później przeprowadzono eksperyment Monte Carlo, porównujący rozkłady miary GMOS i indeksu Jaccarda.

4.5.1. Porównanie własności miar

Poniżej przedstawiono listę pożądanych własności miary podobieństwa prostokątnych regionów rozpoznania $M(L, R)$. W tabeli 4.4 przedstawiono natomiast porównanie miar zaproponowanych w tym rozdziale oraz wybranych miar znalezionych w literaturze.

- A – Niezależność od skali – własność ta pozwala na bezpośrednie porównanie miary podobieństwa par prostokątów opisujących odległe i par prostokątów opisujących bliskie obiekty.

$$\begin{aligned} L_1 \sim_s L_2 \wedge rot(L_1) = rot(L_2) \wedge R_1 \sim_s R_2 \wedge rot(R_1) = rot(R_2) \\ \wedge d(m(L_1), m(R_1)) = d(m(L_2), m(R_2)) \\ \Rightarrow M(L_1, R_1) \simeq M(L_2, R_2) \end{aligned}$$

- B – Wrażliwość na pozycję / możliwość separacji wpływu – miara bierze pod uwagę pozycję porównywanych prostokątów i istnieje możliwość odzyskania informacji o tym, że ten czynnik miał wpływ na wynik.

$$\begin{aligned} A(L_1, R_1) = 1 \wedge A(L_2, R_2) = 1 \wedge A(L_1, L_2) = 1 \\ \wedge S(L_1, R_1) = 1, S(L_2, R_2) = 1 \wedge rot(L_1) = rot(R_1) \wedge rot(L_2) = rot(R_2) \\ \wedge d(m(L_1), m(R_1)) \neq d(m(L_2), m(R_2)) \\ \Rightarrow M(L_1, R_1) \neq M(L_2, R_2) \end{aligned}$$

- C – Wrażliwość na rozmiar / możliwość separacji wpływu – miara bierze pod uwagę rozmiar porównywanych prostokątów i istnieje możliwość odzyskania infor-

macji o tym, że ten czynnik miał wpływ na wynik.

$$\begin{aligned}
 & A(L_1, R_1) \neq A(L_2, R_2) \\
 \wedge & S(L_1, R_1) = 1, S(L_2, R_2) = 1 \wedge \text{rot}(L_1) = \text{rot}(R_1) \wedge \text{rot}(L_2) = \text{rot}(R_2) \\
 & \wedge d(m(L_1), m(R_1)) = d(m(L_2), m(R_2)) \\
 \Rightarrow & M(L_1, R_1) \neq M(L_2, R_2)
 \end{aligned}$$

- D – Wrażliwość na kształt / możliwość separacji wpływu – miara bierze pod uwagę pozycję porównywanych prostokątów i istnieje możliwość odzyskania informacji o tym, że ten czynnik miał wpływ na wynik.

$$\begin{aligned}
 & A(L_1, R_1) = 1 \wedge A(L_2, R_2) = 1 \wedge A(L_1, L_2) = 1 \\
 \wedge & S(L_1, R_1) \neq S(L_2, R_2) \wedge \text{rot}(L_1) = \text{rot}(R_1) \wedge \text{rot}(L_2) = \text{rot}(R_2) \\
 & \wedge d(m(L_1), m(R_1)) = d(m(L_2), m(R_2)) \\
 \Rightarrow & M(L_1, R_1) \neq M(L_2, R_2)
 \end{aligned}$$

- E – Wrażliwość na rotację / możliwość separacji wpływu – miara bierze pod uwagę rotację porównywanych prostokątów i istnieje możliwość odzyskania informacji o tym, że ten czynnik miał wpływ na wynik.

$$\begin{aligned}
 & A(L_1, R_1) = 1 \wedge A(L_2, R_2) = 1 \wedge A(L_1, L_2) = 1 \\
 \wedge & S(L_1, R_1) = 1, S(L_2, R_2) = 1 \wedge \text{rot}(L_1) - \text{rot}(R_1) \neq \text{rot}(L_2) - \text{rot}(R_2) \\
 & \wedge d(m(L_1), m(R_1)) = d(m(L_2), m(R_2)) \\
 \Rightarrow & M(L_1, R_1) \neq M(L_2, R_2)
 \end{aligned}$$

- F – Możliwość nietrywialnego porównania odseparowanych prostokątów – oznacza, że miara jest w stanie dostarczyć informacji nawet w przypadku gdy prostokąty nie mają wspólnych pikseli.

$$\begin{aligned}
 & L_1 \cap R_1 = 0 \wedge L_2 \cap R_2 = 0 \wedge L_1 = L_2 \wedge R_1 \neq R_2 \\
 \Rightarrow & M(L_1, R_1) \neq M(L_2, R_2)
 \end{aligned}$$

- G – Adaptacja wrażliwości na wskazane segmenty prostokąta – funkcja ta reprezen-

Tabela 4.4. Porównanie funkcjonalności miar podobieństwa prostokątnych regionów rozpoznania

Nazwa miary	A	B	C	D	E	F	G
Hausdorff [140]	-	+/-	+/-	+/-	+/-	+	-
RobLoc [141]	+	+/+	-	-	-	+	-
RobCor [141]	+	-	-	-	-	+	-
RobCom [141]	+	-	+	-	-	+	-
FOM [142]	+	+/-	+/-	+/-	+/-	+	-
Hafiane [57]	+	+/-	+/-	+/-	+/-	-	-
Jaccard (3.5)	+	+/-	+/-	+/-	+/-	+	-
Kształt (4.3)	+	-	-	+/+	-	+	-
Pole (4.2)	+	-	+/+	-	-	+	-
Pozycja (4.6)	+	+/+	-	-	-	+	+
Prędkość (4.20)	+	-/-	-/-	-/-	+/+	+	+
GMOS (4.10)	+	+/+	+/+	+/+	-/+	+	+

tuje zdolność do zdefiniowania pewnych części regionów rozpoznania jako szczególnie ważnych w procesie ewaluacji (szczegóły w sekcji 4.4.1 - MOD i TSR).

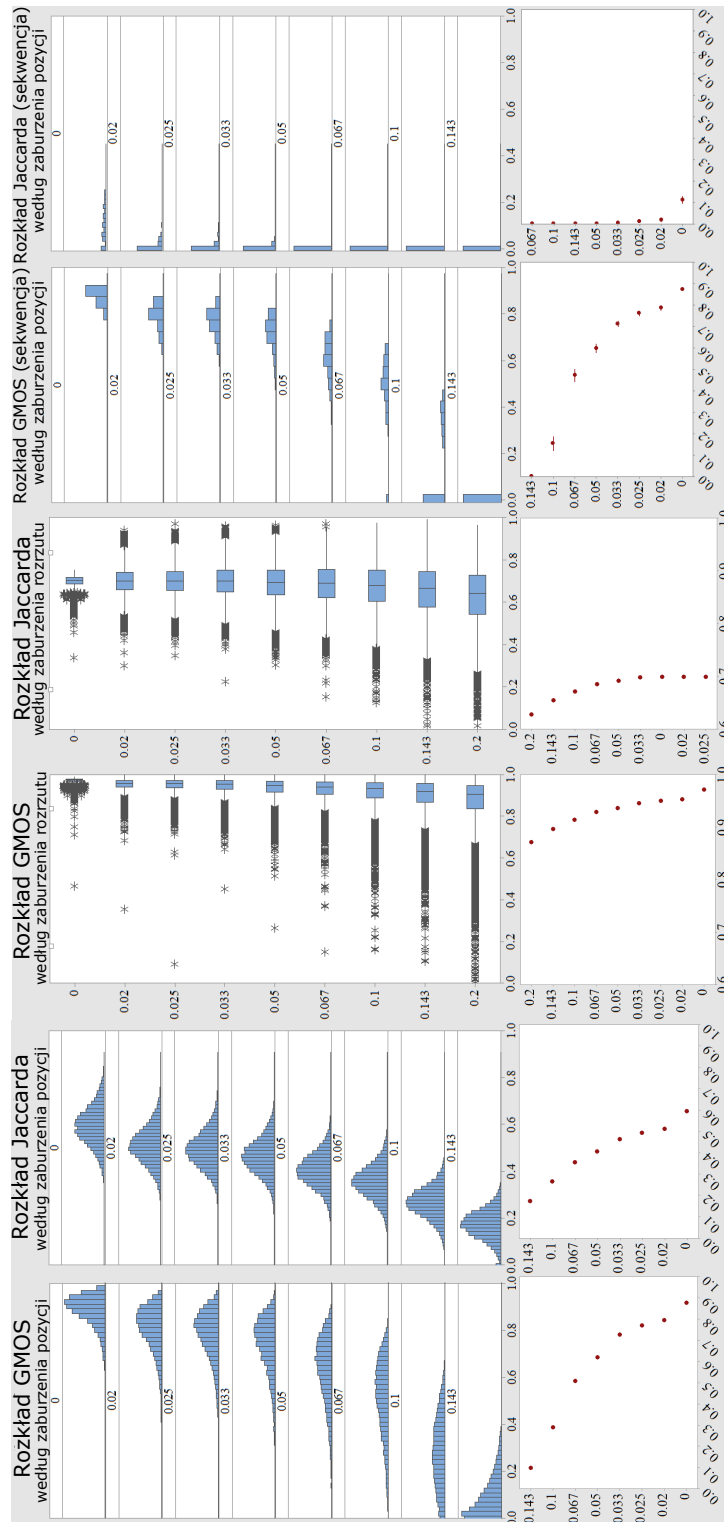
Niech $rot(\cdot)$ to funkcja zwracająca kąt nachylenia prostokąta względem układu odniesienia. Ze względu na fakt, że w pracy zaproponowano miarę podobieństwa prostokątnych regionów rozpoznania oceniającą osobno podobieństwo pola, kształtu i położenia, własności B, C, D są w sposób naturalny spełnione. Oczywiście nie przez wszystkie miary składowe z osobna, ale przez ich kombinację w postaci miary GMOS. Miara służąca do oceny prędkości mogłaby zostać zaimplementowana jako część średniej harmonicznnej ale uznano, że powinna być potraktowana osobno. Powodem jest fakt, że kierunek został ściśle związany z prędkością obiektu i ze względu na analizę w kontekście scenariuszy drogowych traktujemy tę własność w szczególny sposób. Miary zaproponowane w pracy [141] bazujące na podobnej idei separacji własności geometrycznych prostokąta nie uzyskały pełni tego efektu. Warto zauważyć, że indeks Jaccarda nie posiada własności F w swojej podstawowej postaci, a dopiero po zastosowaniu rozszerzenia definicji, tak jak np. zaproponowano to w pracy [81]. Rozszerzenie to jest podobne do uproszczonej wersji, zaproponowanej w pracy, idei miary podobieństwa pozycji. Według tabeli miara Hausdorffa mogłaby stanowić ciekawą propozycję, jednak brak własności A bardzo ogranicza jej użyteczność.

4.5.2. Eksperyment Monte Carlo

Przeprowadzony eksperyment opiera się na autorskim rozszerzeniu metody porównania miar podobieństwa prostokątnych regionów rozpoznania, którą to metodę przedstawiono oryginalnie w pracach [58] i [57]. Został on przeprowadzony, tak aby zbadać zmienność wartości miar zaproponowanych w tym rozdziale z wykorzystaniem wygenerowanych sztucznie wyników o kontrolowanym rozkładzie zaszumienia. W serii eksperymentów wykorzystano ponad 12000 referencyjnych prostokątów pochodzących z laboratorium do etykietowania danych rzeczywistych. Idea eksperymentu polega na wygenerowaniu dla każdego prostokąta z tego zbioru sztucznej odpowiedzi (prostokąta), który zostanie porównany do jego referencji z wykorzystaniem miar podobieństwa prostokątów. Zaburzenie polega na przesunięciu pozycji prostokąta o ustaloną wartość uzależnioną jedynie od jego skali oraz dołożeniu do takiego przesunięcia zaszumienia współrzędnych rogów prostokąta z wykorzystaniem rozkładu normalnego. Eksperyment ten został przeprowadzony na trzy różne sposoby. W pierwszym podejściu analizowano jakość rozpoznania w klatce i możliwość rozpoznania zwiększającego się przesunięcia środka przy ustalonym typie zaszumienia wierzchołków. W drugim przeanalizowano znów jakość rozpoznania w klatce i możliwość rozpoznania zaburzonego wyniku przy zwiększającym się standardowym odchyleniu zaszumienia wierzchołków. W ostatnim eksperymencie analizowano wyniki jakości dla miar sekwencji przy ponownie zwiększającym się przesunięciu i stałych parametrach zaszumienia. Wszystkie wyniki można znaleźć w kolejnych kolumnach na rysunku 4.13. W każdej kolumnie widać na górze rozkłady jakości lub wykresy pudełkowe tych rozkładów w przypadku drugiego eksperymentu. Rozkłady dotyczą uzyskanych wyników jakości dla zestawu referencji i odpowiednio spreparowanych prostokątów. Poniżej można zobaczyć średnie wartości jakości dla wszystkich grup i ich 95% przedziały ufności.

Wpływ przyrostu zmiany pozycji na miary podobieństwa lokalnego przy ustalonym zaszumieniu

Pierwsze dwie kolumny na rysunku 4.13 to wyniki pomiaru jakości w klatce z wykorzystaniem miary podobieństwa GMOS oraz indeksu Jaccarda dla danych o przesuniętym środku ciężkości o odpowiednio: 0, 0,02, 0,025, 0,033, 0,05, 0,067, 0,1, 0,143 długości przekątnej prostokąta referencyjnego w kolejnych próbkach. Tak uzyskany wynik poddano zaszumieniu współrzędnych wierzchołków z wykorzystaniem rozkładu normalnego o parametrach $N(0,05, 0,05)$. Początkowa wartość miary Jaccarda zaczyna się od 0,6 bez przesuwania środka ciężkości (wyłącznie zaszumienie rogów) i spada, gdy przesunięcie ro-



Rysunek 4.13. Wyniki eksperymentu Monte Carlo. Od lewej: porównanie rozkładów ewaluacji w klatce przy zaburzeniu pozycji, przy zaburzeniu rozrzutu oraz ewaluacja sekwencji przy zaburzeniu pozycji

śnie bez znaczącej zmiany kształtu histogramu wartości. Generalna miara podobieństwa GMOS dla zerowej wartości przesunięcia środka ciężkości tworzy rozkład wyśrodkowany w okolicy wartości 0,9 i spada, gdy przemieszczenie środka rośnie wraz ze znaczną zmianą kształtu histogramu wartości. W rezultacie miara ogólna utrzymuje wysokie wartości dla szerszego zakresu wartości przemieszczeń — zgodnie z potrzebami małe przesunięcia nie są uważane za znaczącą zmianę. Z drugiej strony, dla dużych wartości przemieszczeń ogólna miara jest znacznie bliższa 0 niż wskaźnik Jaccarda.

Wpływ przyrostu rozrzutu zaszumienia wierzchołków na miarę podobieństwa lokalnego

Następnie podjęto się manipulacji rozrzutem zaszumienia wyłącznie dla współrzędnych pozycji wierzchołków prostokąta, czyli zmieniano wartość parametru σ rozkładu normalnego $N(0,05, \sigma)$. Przyjmował on znów kolejno wartości 0,02, 0,025, 0,033, 0,05, 0,067, 0,1 i 0,143. Wyniki są widoczne w dwóch środkowych kolumnach rysunku 4.13. Środek ciężkości histogramu miary GMOS dla przesunięć narożnika prostokąta wygenerowanego z rozkładem normalnym $\sigma = 0,02$, który uznaliśmy za mały, jest bliski 100% i spada wraz z bardziej chaotycznym przesunięciem narożników. W tych samych warunkach średnia wartość indeksu Jaccarda waha się chaotycznie w przedziale między 0,6 i 0,7 dla różnych wartości σ . Świadczy to o relatywnie dużej podatności na mało istotne zaburzenia, o niemonotonicznej naturze, co dodatkowo może utrudniać poprawną interpretację wyników.

Wpływ przyrostu zmiany pozycji na miarę podobieństwa sekwencji

Ostatecznie podjęto się porównania jakości sekwencji dla wskazanych dwóch miar wartości lokalnej przy założeniu standardowego podejścia do podsumowania całego zdarzenia (średnia ocen w klatkach). Założono też, że progiem, dla którego uznaje się rozpoznanie dla wartości miary Jaccarda jest 0,7, tak jak to standardowo jest przyjmowane podczas analiz z jego wykorzystaniem. Mapowanie dla miary GMOS jest opisane w sekcji 4.1.5. W przypadku tego eksperymentu zasady generowania wyników były identyczne jak w pierwszym eksperymencie, a wyniki są zawarte w ostatnich dwóch kolumnach rysunku 4.13. Dla przesunięcia równego 0, wartości miary GMOS pozostają powyżej 0,9 i maleją wraz ze wzrostem przesunięcia środka ciężkości. Indeks Jaccarda jest wyraźnie wprowadzany w błąd przez przesunięcie rogów pola, połączone z podejściem progowym dla akceptacji wyniku. Nawet dla braku przesunięcia środka ciężkości prostokąta, wyniki są rzadko uznawane za warte porównania, co mocno obciąża średnią dla całej sekwencji.

Wraz ze wzrostem przesunięcia środka ciężkości wartość oceny zdarzenia szybko spada do 0.

Rozdział 5

Meta-analiza jakości dla zbiorów scenariuszy testowych

5.1. Dane i opis eksperymentu

Opisane w poprzednim rozdziale metryki będą potrzebne do przeprowadzenia eksperymentu, polegającego na analizie jakości działania modułu percepcji MOD, w różnych warunkach otoczenia. Ponadto analizowana jest odporność tego modułu na zaburzenia danych pochodzących z sensora (kamery frontowej). Niniejszy eksperyment ilustruje typowe problemy napotymane podczas procesu testowania modułu percepcji. Możliwość wtórnej analizy jakości na już zebranych i oznaczonych danych testowych pozwala poszerzyć wiedzę na temat zachowania algorytmów wizyjnych w warunkach, które nie występowały na oryginalnych nagraniach przez manipulację danymi pochodzącymi z sensora. Aby osiągnąć postawione cele potrzebna jest nie tylko metodologia obliczania jakości detekcji w scenach, ale też metodologia analizy dużych zbiorów danych rozróżnialnych ze względu na różne i często nakładające się zakłócenia. Zakłócenia te można podzielić na:

- intencjonalnie zaaplikowane i różniące się dalej co do sposobu degradacji jakości oraz poziomu jej nasilenia,
- naturalne, wynikające z warunków jakie panowały na drodze podczas nagrania oraz stosunku sygnał-szum charakteryzującego użyte sensory.

W tym rozdziale podjęta zostaje próba usystematyzowania tak szeroko zakrojonej i wie-

lopoziomowej analizy jakości, która stanowi jedno z kluczowych wyzwań w dziedzinie testowania samochodów autonomicznych.

5.1.1. Tworzenie układu eksperymentalnego

Dane

Do eksperymentu wybrano 108 scen ze zbioru nagrań samochodów testowych Aptiv, które posiadały dane referencyjne dla modułu percepcji MOD. Klatki analizowane są w skali szarości o 12 bitowej głębi oraz mają rozdzielczość 1280×971 pikseli. Jedna scena zawiera około 1800 klatek wideo co odpowiada nagraniu o długości 1 minuty z kamery nagrywającej 30 klatek na sekundę. Średnia liczba obiektów uwzględniając pełny czas ich obecności w jednym nagraniu to 14.

Podział danych ze względu na warunki

Porównanie jakości rozpoznania, w zależności od warunków na scenie, wymaga podzielenia zbioru scen ze względu na dominujący czynnik, mający wpływ na skuteczność działania modułu percepcji. Na wybór kluczowych czynników miała też wpływ możliwość wprowadzenia do obrazu sztucznych zakłóceń w późniejszych etapach eksperymentu. Ostatecznie wydzielono 6 rodzajów scen widocznych na rysunku 5.1. Każdej scenie został przypisany identyfikator GX , gdzie X oznacza typ grupy. Do automatyzacji podziału wszystkich danych wykorzystano informacje zawarte w danych referencyjnych tych scen.

Grupowanie scen według oryginalnej jakości

Podział na grupy pozwala zbadać wpływ zaburzeń na jakość detekcji w różnych warunkach oryginalnie charakteryzujących scenę. Kolejnym krokiem jest podzielenie każdej grupy na podgrupy, uwzględniające kształt rozkładu jakości detekcji osiągnięty przez badany algorytm w scenie bez sztucznych zniekształceń. Pomiar jakości rozpoznania obiektów w scenie przeprowadzono z wykorzystaniem metody zaproponowanej w rozdziale 4. Oznacza to, że dla każdej sceny zawierającej n obiektów, opisanych w danych referencyjnych sekwencjami prostokątnych regionów rozpoznania, istnieje możliwość przypisania zbioru składającego się z n liczb, będących wartościami miary podobieństwa sekwencji (4.12).



Rysunek 5.1. Sześć grup scenariuszy testowych wydzielonych na podstawie warunków dominujących podczas sceny oraz ich symbole

5.1.2. Algorytm klastrowania jakościowego scen

Podział grup scen na podgrupy, w oparciu o jakość detekcji, ma na celu zbadanie wpływu symulowanych zakłóceń na degradację skuteczności wykrywania i śledzenia obiektów przez badany algorytm na zbiorach, dla których skuteczność jego działania była podobna. Aby przygotować taki układ eksperymentalny wykorzystano algorytm klastrowania K-średnich, opisany w [143]. Jako pojedynczy obiekt uznaje się empiryczny rozkład jakości w danej scenie μ . Niech zbiór wartości miar podobieństwa sekwencji obiektów w jednej scenie to q_1, \dots, q_n ($n \in \mathbb{N}, q_i \in (0, 1]$), a przyporządkowanym prawdopodobieństwem empirycznym wystąpienia danej obserwacji w tej scenie jest $\mu(q_1) = \mu(q_2) = \dots = \mu(q_n) = \frac{1}{n}$, co oznacza, że

$$\sum_{i=1}^n \mu(q_i) = 1. \quad (5.1)$$

Ponadto miarą odległości między obiektami będzie metryka Wassersteina (patrz podrozdział 3.3.4). Aby obliczyć jej wartość pomiędzy scenami reprezentowanymi przez ich empiryczne rozkłady jakości μ i ν (o zbiorze obserwacji $\{q'_1, \dots, q'_m\}$), należy znaleźć funkcję

$f^* : \{q_1, \dots, q_n\} \times \{q'_1, \dots, q'_m\} \rightarrow [0, 1]$, która minimalizuje sumę:

$$\sum_{k=1}^n \sum_{j=1}^m f(k, j) d(k, j), \quad (5.2)$$

gdzie $d(k, j) = |q_k - q'_j|$. Kryteria dla funkcji f są następujące:

$$\forall k \in \{1, \dots, n\} \forall j \in \{1, \dots, m\} : 0 \leq f(k, j) \leq 1, \quad (5.3)$$

$$\forall k \in \{1, \dots, n\} : \sum_{j=1}^m f(k, j) = \mu(q_k), \quad (5.4)$$

$$\forall j \in \{1, \dots, m\} : \sum_{k=1}^n f(k, j) = \nu(q'_j), \quad (5.5)$$

$$\sum_{k=1}^n \sum_{j=1}^m f(k, j) = 1. \quad (5.6)$$

Aby rozwiązać taki układ należy zastosować metodę sympleks, a dokładniej jej odmianę przeznaczoną dla sieci, opisaną między innymi w [144] i [145]. Następnie wartość odległości Wassersteina może być wyliczona ze wzoru:

$$W(\mu, \nu) = \sum_{k=1}^n \sum_{j=1}^m f^*(k, j) d(k, j). \quad (5.7)$$

Centroidy i inercja w kontekście rozkładów empirycznych

Zakłada się, że k jest liczbą podgrup na jakie zamierzono podzielić daną grupę scen testowych, w wyniku zastosowania algorytmu klastrowania. N jest liczbą zawartych w tej grupie scen testowych. Niech μ_i dla $i = 1, \dots, N$ to empiryczne rozkłady jakości scen należących do tej grupy. Empiryczny rozkład jakości jest dyskretnym rozkładem prawdopodobieństwa, uzyskanym przez normalizację rozkładu miar jakości rozpoznania obiektów w scenie. Celem algorytmu jest minimalizacja inercji przyjmującej następującą formę:

$$I = \sum_{j=1}^k \sum_{\mu_i \in C_j} W(\mu_i, \xi_j)^2. \quad (5.8)$$

We wzorze tym C_j dla $j = 1, \dots, k$ są to klastry scen, a ξ_j są centroidami tych klastrów. Dla każdego kroku algorytmu poza krokiem pierwszym (inicjującym), centroidy to rozkłady prawdopodobieństwa uśredniające, w sensie metryki Wassersteina, wszystkie rozkłady

przynależące do danego klastra. Taki konstrukt nazywa się barycentrum Wassersteina [130]:

$$\xi_j = \arg \min_{\xi \in P(\mathcal{M})} \frac{1}{|C_j|} \sum_{\mu \in C_j} W(\xi, \mu). \quad (5.9)$$

Wybór centroidów inicjujących przeprowadzono metodą podobną do metody zaproponowanej w pracy [146]. Dla $i = 1, \dots, N$ definiuje się:

$$D(\mu_i, \Xi) = \min\{W(\mu_i, \xi_j) : j = 1, \dots, |\Xi|\}, \quad (5.10)$$

gdzie Ξ jest zbiorem wszystkich centroidów inicjujących dotychczas wybranych. $|\Xi| \leq k$. Pierwszym centroidem ξ_1 zostaje arbitralnie wybrany rozkład μ_i ze sceny o największej liczbie obserwacji. Każda następna scena jest dołączona pod warunkiem, że ma najwyższe prawdopodobieństwo zostania kolejnym centroidem inicjującym:

$$G(\mu_i) = \frac{D(\mu_i, \Xi)^2}{\sum_{t=1}^N D(\mu_t, \Xi)^2}. \quad (5.11)$$

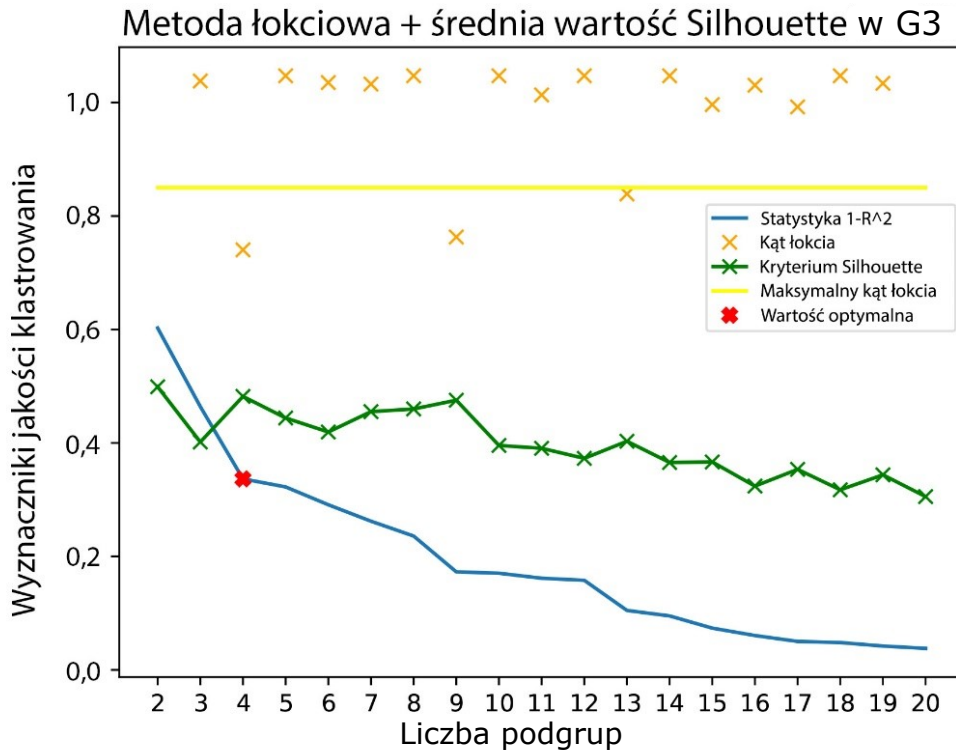
Funkcja prawdopodobieństwa G jest przeliczana od nowa, po każdej aktualizacji zbioru Ξ .

Wybór optymalnej liczby podgrup

Aby określić, na jaką optymalną liczbę podgrup k należy podzielić daną grupę rozkładów oryginalnej jakości, zdecydowano o połączeniu dwóch metod. Pierwszą jest metoda łokciowa dla statystyki $(1 - R^2)$ nazywanej procentem wariancji niewyjaśnionej. Zakładając, że po zastosowaniu algorytmu klastrowania, a $\Xi' = \{\xi'_1, \dots, \xi'_k\}$ jest zbiorem znalezionych w jego konsekwencji centroidów dla $k \in \{2, \dots, N\}$. Φ jest barycentrum Wassersteina obliczonym dla wszystkich rozkładów przeznaczonych do tego klastrowania. Wtedy można wyznaczyć statystykę:

$$1 - R^2 = 1 - \frac{\sum_{j=1}^k W(\xi'_j, \Phi)^2}{\sum_{i=1}^N W(\mu_i, \Phi)^2}. \quad (5.12)$$

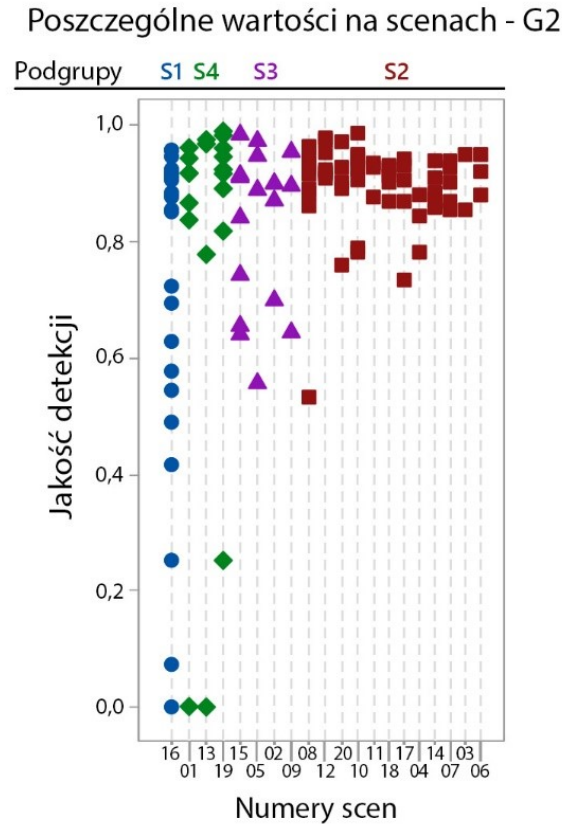
Dla $k = N$ cała wariancja jest wyjaśniona przez model zaproponowany przez algorytm i powyższa statystyka jest równa 0. Za optymalne k , w sensie tej metody, uznaje się liczbę podgrup, przy której spadek procentu wariancji niewyjaśnionej przestaje być in-



Rysunek 5.2. Statystyki związane z procesem podejmowania decyzji o optymalnej liczbie podgrup na jakie zostanie podzielona grupa G3, podczas jej klastrowania. Optimum w tym przypadku zostało wybrane na podstawie metody łokciowej

tensywny. Na wykresie tej statystyki względem liczby podgrup powinien być widoczny tytułowy „łokieć” nad optymalną liczbą podgrup. Motywacją do wykorzystania tej statystyki do wyboru liczby podgrup jest fakt, że w późniejszej fazie eksperymentu zostanie wykorzystana analiza wariancji. Redukcja wpływu czynników, niezwiązanych z głównym przedmiotem eksperymentu, na wariancję, będzie miała dobry wpływ na wyniki pod względem ich czytelności, wiarygodności i mocy testów. Ponieważ zdarza się jednak, że metoda łokciowa nie jest rozstrzygająca, zaproponowano wykorzystanie średniej wartości silhouette [147]. Niech C_1, \dots, C_k to klastry, do których przypisano powyższe centroidy Ξ' . Dla każdego $\mu \in C_j$, $j = 1, \dots, k$ definiuje się:

$$a(\mu) = \frac{\sum_{\mu' \in C_j, \mu' \neq \mu} W(\mu', \mu)}{|C_j|} \quad (5.13)$$



Rysunek 5.3. Klastry scen z G2 wyznaczone przez algorytm. Na poziomej osi znajdują się numery scen, znaczniki reprezentują wartości miary podobieństwa sekwencji obiektów w tych scenach, a ich kolory wyznaczają klastry

i

$$b(\mu) = \min_{\substack{t=1,\dots,k \\ t \neq j}} \left\{ \frac{1}{|C_t|} \sum_{\mu' \in C_t} W(\mu', \mu) \right\} \quad (5.14)$$

wtedy silhouette dla μ to

$$s(\mu) = \frac{b(\mu) - a(\mu)}{\max\{a(\mu), b(\mu)\}}, \quad (5.15)$$

o ile μ nie jest jedynym rozkładem w klastrze, w takim wypadku $s(\mu) = 1$. Liczba podgrup, dla których średnia wartość silhouette dla wszystkich klastrowanych scen jest najwyższa, uważana jest za optymalną w sensie tej metody [148].

Na rysunku 5.2 można zobaczyć proces podejmowania decyzji o optymalnej liczbie podgrup wybranej do klastrowania grupy G3. Oczekiwanym wynikiem klastrowania są podgrupy spójne pod względem średniej jakości rozpoznania, odchylenia standardowego i rozrzutu obserwacji odstających. Tak skonstruowany układ eksperymentalny pozwoli określić na ile oryginalny rozkład miar jakości detekcji może gwarantować potencjalną

odporność modułu percepcji na degradację danych lub pracę w warunkach trudniejszych niż te, na których oryginalnie go testowano. Dodatkowo ujednoczenie rozrzutu w podgrupach uczyni wyniki testów statystycznych bardziej precyzyjnymi. Poniżej przyjęto następujące oznaczenia dla podgrup danych: $GxSy$, gdzie x oznacza numer grupy (analogicznie jak na rysunku 5.1), a y numer podgrupy wydzielonej w danej grupie na podstawie klastrowania.

5.1.3. Rodzaje analizowanych zaburzeń

Aby przeanalizować odporność modułu percepcji na degradację danych z kamery porównana zostanie jakość detekcji modułu MOD pracującego na danych oryginalnych (niezaburzone filmy), a następnie wprowadzony zostanie jeden z czterech wybranych sposobów degradacji obrazu wideo:

- rozmycie Gaussa,
- redukcja głębi bitowej skali szarości,
- okluzja cząstkowa,
- cykliczne oślepienie.

Tak zakłócone dane wizyjne zostaną poddane resymulacji i przetworzeniu przez moduł percepcji, a wyniki jego działania ponownie ocenione z wykorzystaniem tych samych danych referencyjnych. Każda z metod sztucznej deformacji danych posiadać będzie również regularnie podnoszony poziom intensywności, i każdy zestaw tak zdegradowanych danych osobno zostanie poddany procesowi wykrywania obiektów i analizie jego jakości.

Rozmycie gaussowskie

Pierwszym ze sposobów deformacji obrazu jakie postanowiono przebadać jest rozmycie Gaussa. Zostało ono wybrane, ponieważ przy jego pomocy, można reprodukcować wpływ na obraz z kamery deszczu, śniegu, mgły lub utraty ostrości kamery. Degradacja została wprowadzona przez splot dwuwymiarowej macierzy o wartościach reprezentujących gęstość rozkładu normalnego o ustalonych parametrach z podzbiorem pikseli obrazu [149]. Intensywność zaburzenia była sterowana rozmiarem macierzy kwadratowej. Przetestowane zostały rozmiary 3, 9, 15, 21, 27, 33, 39, 45, 51, 57 i 65 pikseli. Im większy rozmiar macierzy tym intensywniejszy poziom rozmycia uzyskano.



Rysunek 5.4. Przykładowy fragment klatki filmu z wprowadzonymi różnymi poziomami rozmycia Gaussa. Od lewej - oryginał i rozmycia na poziomach 9, 27, 45 i 65



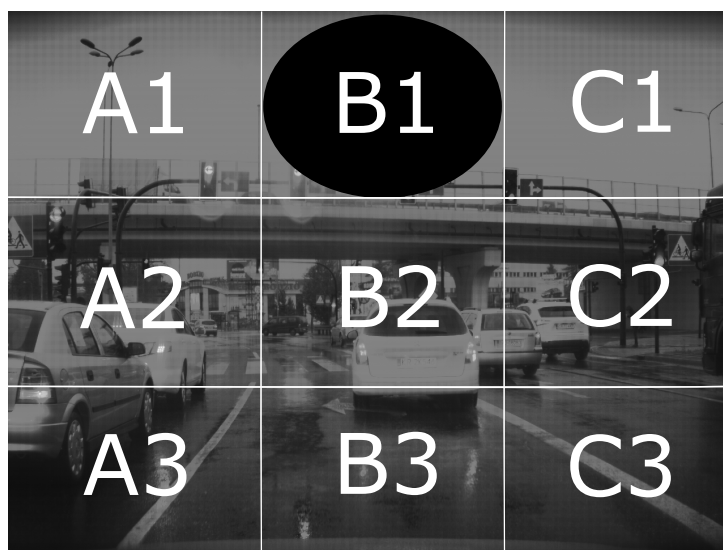
Rysunek 5.5. Przykładowy fragment klatki filmu po wyzerowaniu kolejno 0, 3, 8, 10, 11 ostatnich bitów rozdzielczości skali szarości

Redukcja poziomu szarości

Kolejnym celem, jaki postawiono w tym eksperymencie, jest przeanalizowanie wpływu redukcji głębi bitowej na jakość detekcji obiektów. Obrazy oryginalnie posiadały 12 bitów głębi bitowej w skali szarości, co oznacza 4096 poziomów szarości. Degradacja danych polegała na wyzerowaniu kolejno od 1 do 11 najmniej istotnych bitów każdego piksela. Oznacza to, że w przypadku gdzie wyzerowano piąty bit od końca, wyzerowane również zostały bity od 1 do 4. Odcinanie kolejnych bitów w każdym kroku zmniejszało pulę poziomów szarości o połowę. W ostatnim eksperymencie po redukcji 11 bitów obrazy, na których moduł percepcji poszukiwał obiektów miały tylko 2 poziomy szarości.

Częściowa okluzja obrazu

Następnym przykładem wprowadzonego do obrazu utrudnienia dla modułu percepcji jest częściowe przesłonięcie jednego z dziewięciu fragmentów obrazu przy pomocy elipsy. Ma ona odzwierciedlać naturalne zakłócenie wprowadzone przez pył, kurz lub drobne obiekty, które utknąwszy na soczewce kamery przesłaniają część pola widzenia. Celem było wskazać jak mocno spadnie średnia jakość rozpoznawania obiektów w zależności od części obrazu, która jest zasłonięta.



Rysunek 5.6. Podział klatki na segmenty, które zostały kolejno przesłonięte. W segmencie B1 znajduje się przykładowa elipsa imitująca fizyczne przesłonięcie sensora



Rysunek 5.7. Cykl oślepienia wprowadzony do filmu. Jeden cykl trwa 26 klatek, pierwsze $26 - k - 2s$ klatek jest niezmiennych, następnie w s klatkach poziom oślepienia narasta, przez k klatek jest na maksymalnym poziomie m , po czym przez s klatek spada do zera

Cykliczne oślepienie

Ostatnim badanym sposobem modyfikacji obrazu wideo jest manipulowanie poziomem bieli pikseli, imitujące chwilowe oślepienie sensora. Każdy piksel, w klatce filmu, zostanie poddany intensyfikacji poziomemu. Polega ona na dodaniu ustalonej wartości do poziomu bieli pikseli obrazu. Wprowadzone „oślepienie” będzie cyklicznie pojawiać się w filmie według następującego schematu. Najpierw ustalono poziom maksymalnej intensyfikacji poziomu bieli $m \in [0, 1]$. Każdy cykl składa się z 26 klatek wideo, w których $k \in \mathbb{N}, k < 26$ posiadać będzie maksymalnie zintensyfikowany poziom bieli m . Cykl rozpoczyna się od serii klatek niezmiennych, następnie intensyfikacja bieli pikseli będzie wzrastać liniowo od 0 aż do osiągnięcia maksimum na przestrzeni s klatek, a po k klatkach o maksymalnym natężeniu „oślepienia” natężenie spadnie do zera znów na przestrzeni s klatek, po czym

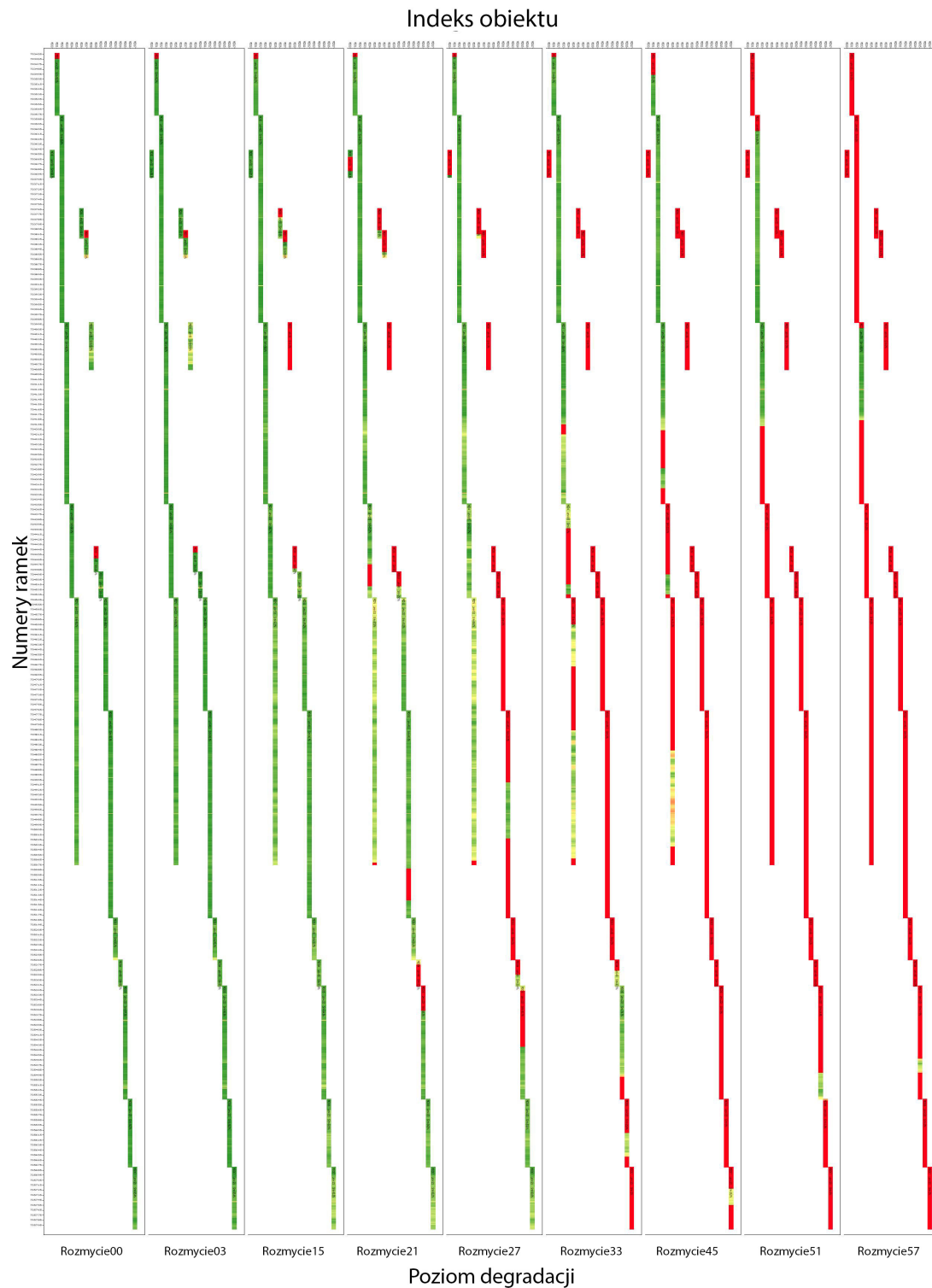
cykl rozpoczyna się od nowa. Zatem zmodyfikowanych klatek w jednym cyklu będzie $k + 2s$, a niezmodyfikowanych $26 - k - 2s$. Typ cyklu wprowadzonego do filmu został zakodowany w następujący sposób: $s - k - m$.

5.2. Wyniki

Wyniki eksperymentów analizowano na kilka sposobów. Surowa degradacja jakości GMOS z wykorzystaniem rozmycia na przykładowej scenie, jest zaprezentowana na rysunku 5.8, z wykorzystaniem metody wizualizacji, zaproponowanej w sekcji 4.3. Można na nim zauważyć, że spadek jakości podobieństwa sekwencji jest powodowany bardziej przez często powtarzający się brak detekcji obiektu, a w mniejszym stopniu przez pogorszenie jakości samej detekcji w poszczególnych klatkach (choć taką też można odnotować). Dla każdej podgrupy, dla każdego typu zaburzenia, na każdym jego poziomie, jakość rozpoznania obiektów w scenach oceniono przy pomocy zdefiniowanej w sekcji 4.2 miary podobieństwa sekwencji. Tak otrzymane zestawy wyników poddano analizie, zgodnie z zasadami testowania hipotez statystycznych (ANOVA). Jednocześnie obliczono odległości Wassersteina tych rozkładów miar jakości od rozkładów uzyskanych dla danych niezaburzonych. Poniżej, w podsekcjach omówiono wyniki tej analizy dla wybranych grup i podgrup scen testowych. Dokładniej omówiono przykłady grup i podgrup gdzie uzyskano wyniki nietypowe oraz podjęto próbę ustalenia przyczyn takich efektów. Podsumowanie całego eksperymentu i konkluzje znajdują się w tabelach na końcu sekcji.

5.2.1. ANOVA w podgrupach

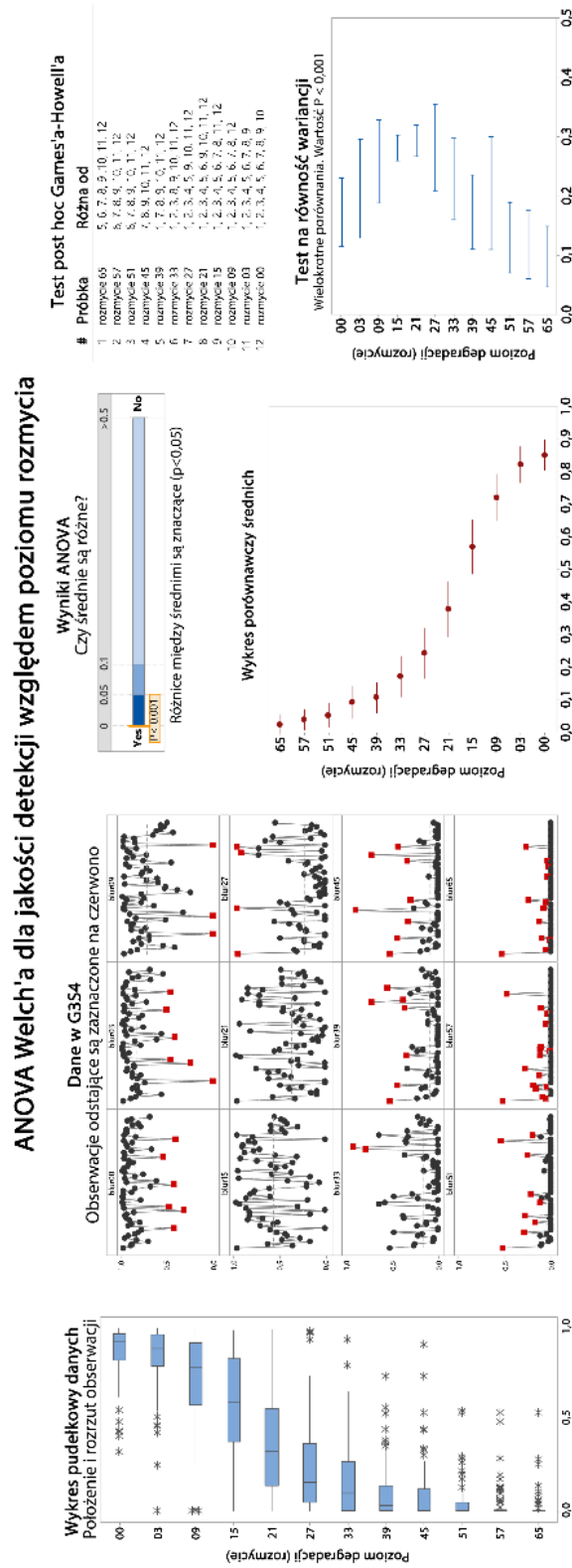
W celu przestudiowania degradacji na podstawie dwóch głównych parametrów rozkładów, wykorzystano analizę wariancji (ANOVA) z poprawką Welch'a, ponieważ utrzymuje ona błąd pierwszego rodzaju alfa na stałym poziomie lepiej, niż jej podstawowa wersja lub testy nieparametryczne w warunkach nierównych liczebności podgrup o różnych wariancjach [150]. Poziom istotności testów wynosił 5%. Obliczenia wykonano z wykorzystaniem oprogramowania statystycznego Minitab [151], który udostępnia ten i pozostałe, wykorzystane poniżej rodzaje testów. O istotnej różnicy między średnimi informują, przeprowadzone dodatkowo, testy post hoc Gamesa Howella. Test Gamesa-Howella to najlepszy test, który zakłada poprawkę na nierówne wariancje w porównywanych próbkach, a także na nierówne liczebności prób [152]. Procedura ta jest używana zaraz po wykonaniu analizy wariancji Welcha, w sytuacji, gdy w wyniku testu Levene'a [153], zostanie od-



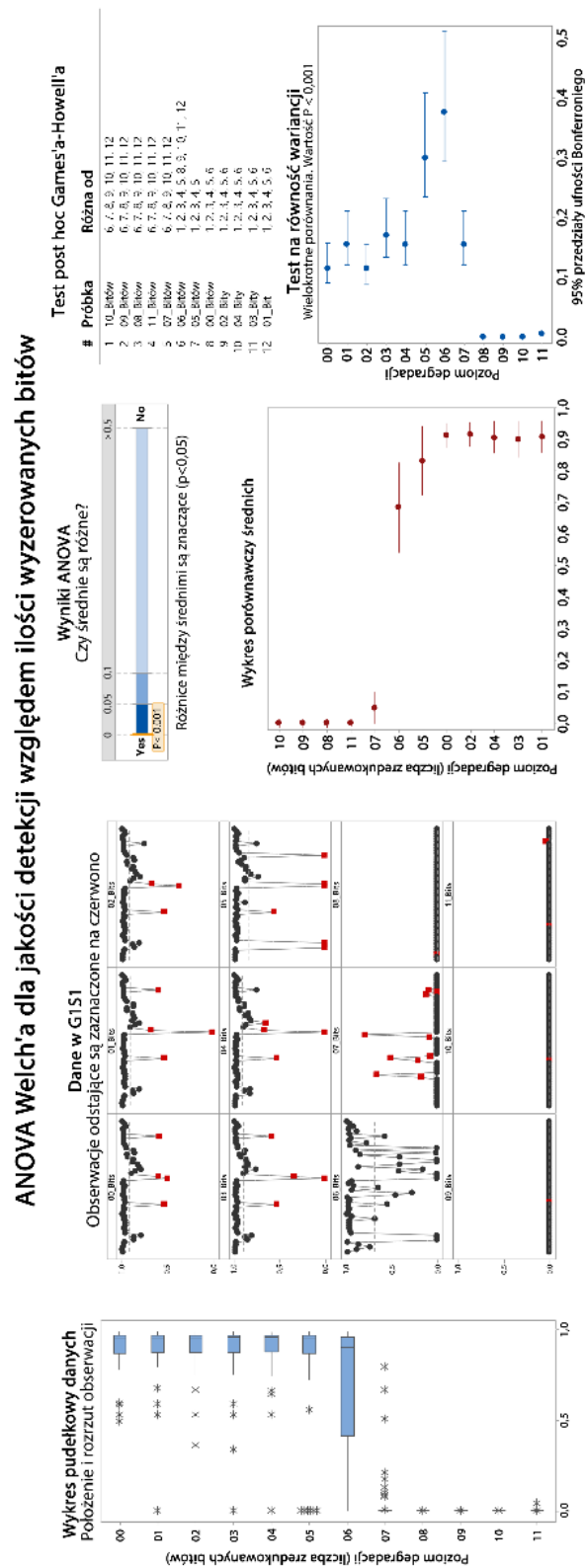
Rysunek 5.8. Wizualizacja degradacji jakości GMOS (4.10) w klatkach przykładowego filmu. Na kolejnych segmentach widzimy tę samą scenę analizowaną przez moduł percepcji przy pogłębiającym się poziomie degradacji obrazu (rozmycie Gaussa)

rzucana hipoteza zerowa o homogenicznych wariancjach w próbach. Przykładowe wyniki tych testów w wybranych podgrupach odpowiednio dla rozmycia, redukcji skali szarości, okluzji i oślepienia można znaleźć na rysunkach 5.9, 5.10, 5.11 i 5.12. Przedstawiono na nich wykresy pudełkowe danych, wykresy punktowe wszystkich obserwacji w wybranej podgrupie, p-value dla testu ANOVA, średnie próbek oraz ich przedziały ufności dla wszystkich poziomów danego zaburzenia, wyniki testu Gamesa Howella oraz przedziały ufności dla wariancji wizualizujące wyniki testowania równości wariancji w próbkach.

Panel na rysunku 5.9 przedstawia analizę parametrów w podgrupie G3S4. Można z niego odczytać, że istotną statystycznie różnicę potwierdzono w tym wypadku już dla rozmycia macierzą o rozmiarze 9. Spadek jakości jest stopniowy wraz ze wzrostem rozmiaru macierzy, dla rozmiarów pośrednich (15,21,27) przedziały ufności dla średniej są tylko nieznacznie szersze, jakość jest nieistotnie różna od zera dla rozmiaru 57 i większego. Tabela 5.1 zawiera wartości średniej w każdej podgrupie przy wszystkich poziomach zniekształcenia obrazu. Na czerwono zaznaczono średnie, dla których na podstawie analizy wariancji udało się potwierdzić statystyczną istotność różnicy względem średniej próbki bez degradacji. Na niebiesko zaznaczono średnie próbek, dla których nie ma podstaw do odrzucenia hipotezy zerowej o równości średnich względem próbki bez sztucznego rozmycia. Generalnie można stwierdzić, że przy rozmiarze macierzy rozmycia przekraczającym 9 pikseli należy oczekiwać istotnej degradacji jakości rozpoznania, wyjątki pojawiają się dość regularnie, co omówiono dalej. Panel na rysunku 5.10 przedstawia analizę parametrów rozkładu jakości w podgrupie G1S1 dla powiększającej się liczby wyzerowanych, końcowych bitów kodujących rozdzielczość dostępnych poziomów szarości. Po wyzerowaniu nawet pięciu ostatnich bitów nie następuje istotna statystycznie utrata średniej jakości rozpoznawanych obiektów, ale przy wyzerowaniu 5 bitów wariancja próbki staje się istotnie statystycznie większa niż dla próbki o wyzerowanych 4 bitach, która z kolei nie ma istotnie różnej wariancji od oryginalnej jakości. Na panelach przedstawiających pojedyncze obserwacje widać, że dla większości obiektów pogorszenie jakości, jeśli nastąpiło to oznaczało natychmiastowy spadek do 0%. Po wyzerowaniu 7 bita następuje najbardziej gwałtowny spadek średniej jakości, wyzerowanie kolejnych bitów sprawia, że średnia próbki przestaje być statystycznie istotnie różna od zera. W tabeli 5.2 przedstawiono podsumowanie wyników analizy wariancji dla wszystkich podgrup. Można z niej odczytać, że generalnie udaje się potwierdzić istotność różnicy w średniej względem oryginalnych danych po wyzerowaniu bitów od szóstego wzwyż. Bardziej odporne przypadki podgrup zostaną omówione później. Dla jednej z podgrup, po wyzerowaniu bitów od piątego wzwyż, można potwierdzić istotność statystyczną, ale ponieważ błąd alfa dla testu

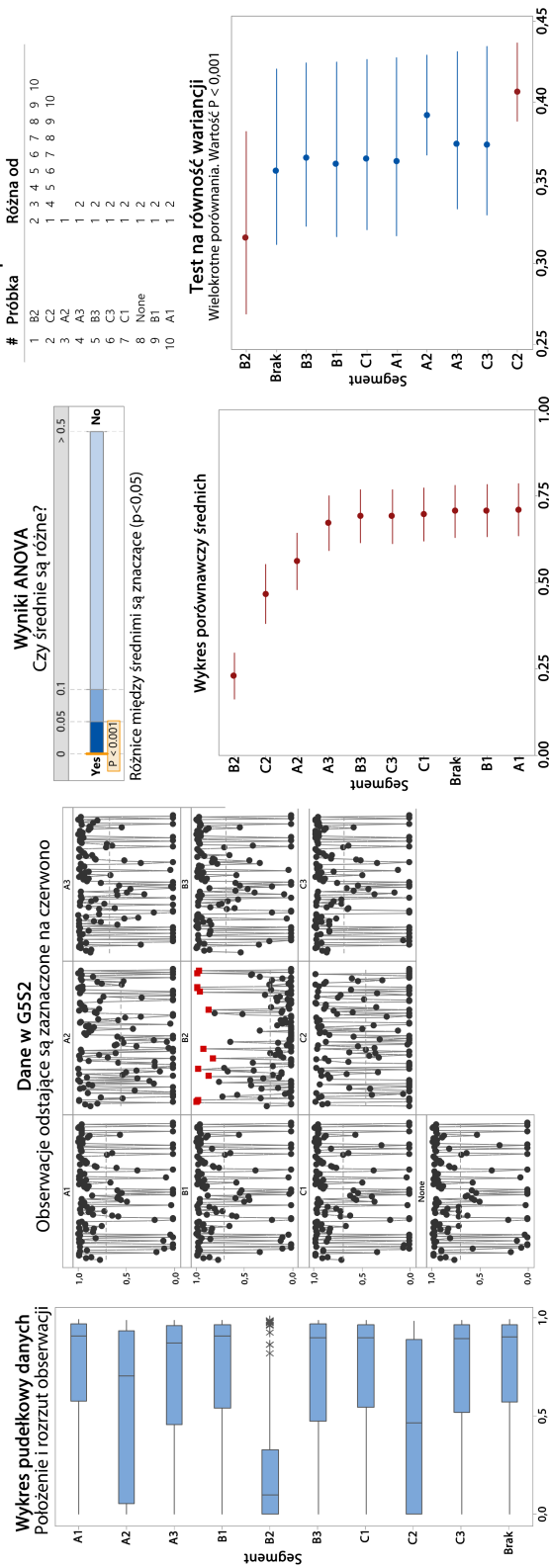


Rysunek 5.9. Analiza wariancji w podgrupie G3S4 dla różnych rozkładów jakości rozpoznawania obiektów przy zwiększającym się poziomie rozmycia Gaussa

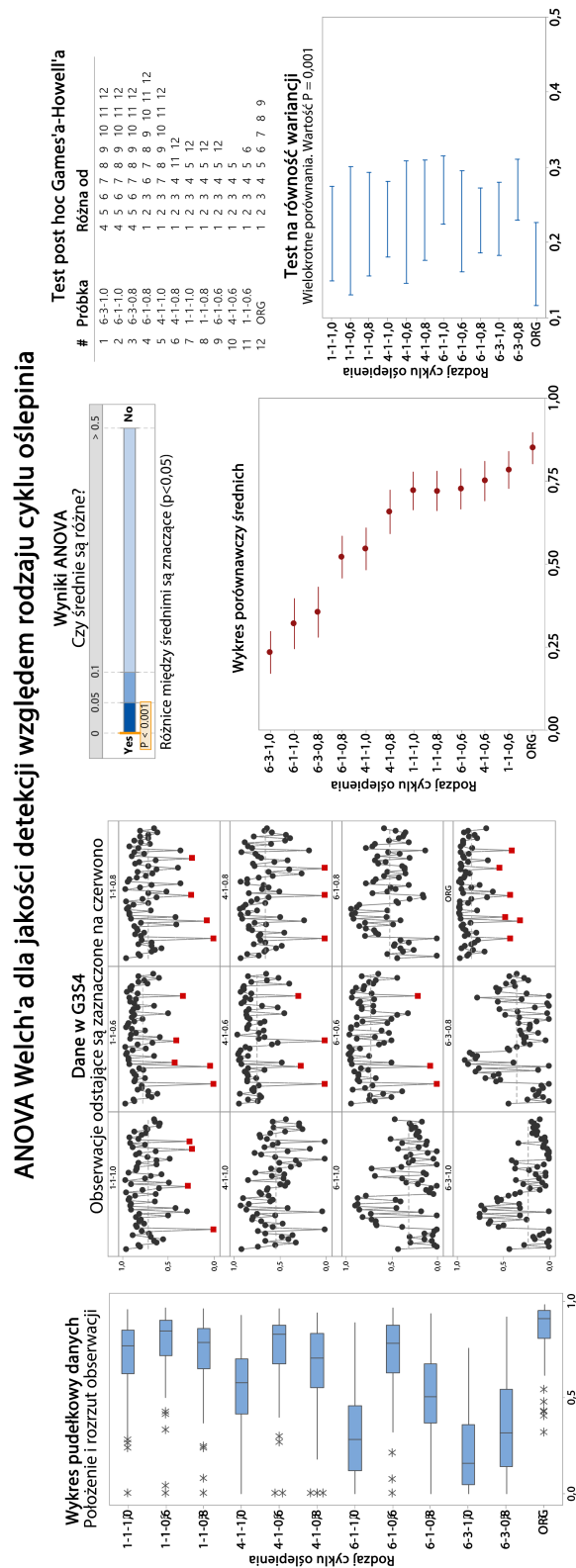


Rysunek 5.10. Analiza wariancji w podgrupie G1S1 dla różnych rozkładów jakości rozpoznawania obiektów przy zmniejszającej się rozdzielczości poziomów szarości

ANOVA Welch'a dla jakości detekcji względem zakrytego segmentu



Rysunek 5.11. Analiza wariancji w podgrupie G5S3 dla różnych rozkładów jakości rozpoznawania obiektów przy okluzji obrazu obejmującej dziewięć różnych segmentów



Rysunek 5.12. Analiza wariancji w podgrupie G3S4 dla różnych rozkładów jakości rozpoznawania obiektów przy odmiennych typach cyklicznego oślepienia

ANOVA wynosi 5%, wskazanym jest to odstępstwo zignorować. Aby wzmocnić konkluzję o nieistotności wpływu zerowania bitów na piątej pozycji i niższych, opierając się na statystycznym testowaniu hipotez, potrzeba więcej danych do osiągnięcia wystarczającego poziomu mocy testów, który w aktualnym eksperymencie wyniósł około 70% w zależności od grupy. Aby przeanalizować dogłębniej wyniki (szczególnie dane dotyczące bitów na pozycjach bardziej znaczących niż 5), wyniki analizy wariancji zostaną uzupełnione o dodatkową analizę topologiczną w kolejnej sekcji.

Panel na rysunku 5.11 przedstawia analizę parametrów w podgrupie G5S3 dla sztucznej okluzji fragmentu obrazu. Istotną statystycznie różnicę w średniej jakości rozpoznania udało się potwierdzić przy zasłonięciu segmentów B2 i C2. Mimo widocznego spadku w przypadku segmentu A2, nie ma podstaw do odrzucenia hipotezy zerowej o równości średnich w tej podgrupie, jednak po uwzględnieniu wyników dla wszystkich podgrup przedstawionych w tabeli 5.3 można stwierdzić, że wszystkie trzy segmenty znajdujące się w rzędzie drugim powodują istotną statystycznie degradację średniej jakości rozpoznania.

Panel na rysunku 5.12 przedstawia analizę parametrów w podgrupie G3S4 w eksperymencie polegającym na cyklicznym sztucznym oślepieniu sensora. Istotną statystycznie różnicę średniej jakości względem rozpoznawania, w warunkach bez wprowadzonej degradacji, można potwierdzić dla wszystkich badanych cykli, oprócz $1 - 1 - 0.6$ i $4 - 1 - 0.6$. Najsłabsze wyniki, istotnie różne w sensie średniej od wszystkich pozostałych próbek rozpoznano dla cykli $6 - 3 - 1.0$, $6 - 1 - 1.0$, $6 - 3 - 0.8$. W tabeli 5.4 podsumowano wyniki ANOVA dla wszystkich podgrup w eksperymencie. Rozkłady jakości degradowane są proporcjonalnie (bez gwałtownych spadków pojedynczych obserwacji) co przekłada się na niższą wariancję próbek niż w innych eksperymentach, ale jest ona trudna do oszacowania, co bardziej wpływa na zbiorczy rozrzut wyników ANOVA. Mimo to istnieją podgrupy, gdzie nawet najslabszy cykl oślepienia wprowadza istotną różnicę. Można zwrócić uwagę, że większy wpływ na pogorszenie jakości miały dłuższe czasy oślepienia (licząc sumę pośrednich klatek i maksymalnie zmodyfikowanych) niż poziom maksymalnej intensyfikacji bieli.

5.2.2. Analiza topologiczna

Aby dokładniej opisać wpływ degradacji obrazu na jakość rozpoznawania obiektów w próbkach eksperymentalnych, odwołano się do innego narzędzia, opisującego różnicę w rozkładach prawdopodobieństwa. Chodzi o odległość Wassersteina, która okazała się już wcześniej użyteczna podczas klastrowania oryginalnych rozkładów jakości rozpoznania w filmach bez degradacji. Choć jest to narzędzie służące bardziej do analizy topologicznej

Tabela 5.1. Podsumowanie wyników ANOVA dla wprowadzonego rozmycia o różnym natężeniu i korespondujące wartości Wassersteina

Natężenie	3	9	15	21	27	33	39	45	51	57	65
G1S1	90,7	89,7	88,2	83,6	72,1	57,3	43,5	38,0	22,7	19,5	7,9
	,018	,020	,029	,075	,190	,338	,476	,531	,684	,715	,831
G1S2	74,2	65,9	53,1	42,4	35,3	24,1	13,6	8,8	6,8	5,7	4,0
	,088	,170	,298	,405	,476	,589	,694	,742	,762	,772	,789
G1S3	92,2	92,2	89,9	87,1	75,5	63,4	58,7	49,5	41,7	33,4	15,4
	,016	,013	,026	,053	,167	,288	,335	,426	,505	,588	,768
G2S1	65,2	50,6	27,1	21,7	16,6	13,3	10,4	6,4	5,3	4,7	2,5
	,058	,205	,439	,493	,544	,577	,606	,646	,658	,663	,685
G2S2	89,3	84,0	70,3	50,6	36,5	24,8	15,9	6,6	3,8	3,3	1,5
	,006	,057	,194	,391	,532	,649	,738	,831	,859	,864	,882
G2S3	79,5	78,6	75,8	65,9	48,5	39,0	33,1	19,8	11,4	9,1	5,3
	,007	,006	,033	,131	,306	,400	,459	,592	,677	,700	,737
G2S4	77,0	53,4	41,3	27,6	18,5	11,4	9,4	6,6	5,5	5,4	0,0
	,057	,292	,413	,550	,641	,712	,731	,759	,770	,771	,826
G3S1	87,7	87,7	85,1	77,9	45,4	24,8	17,3	12,2	6,5	5,5	6,6
	,004	,014	,029	,100	,425	,631	,706	,757	,814	,824	,814
G3S2	91,5	83,8	71,2	53,1	41,5	31,9	23,9	16,1	10,5	5,9	3,0
	,018	,095	,220	,402	,517	,613	,694	,771	,828	,874	,903
G3S3	78,6	70,0	52,0	37,8	29,3	22,7	15,9	11,9	6,3	3,9	2,7
	,009	,086	,266	,409	,493	,559	,628	,667	,724	,748	,759
G3S4	84,8	81,9	71,8	57,7	37,6	24,2	17,1	10,6	9,1	5,2	3,8
	,030	,130	,280	,472	,606	,677	,742	,757	,796	,810	,824
G4S1	90,7	75,2	58,3	45,3	35,7	29,7	27,3	24,0	21,6	14,9	11,1
	,008	,160	,329	,459	,555	,614	,639	,671	,695	,762	,801
G5S1	58,0	51,1	40,1	30,7	21,9	16,4	13,0	8,8	6,5	5,1	3,5
	,041	,108	,219	,312	,401	,456	,489	,532	,555	,569	,585
G5S2	86,2	80,9	65,6	50,3	34,3	24,5	18,1	13,8	11,1	7,6	4,6
	,018	,066	,219	,372	,533	,630	,694	,737	,764	,799	,829
G5S3	72,5	63,7	52,9	41,3	30,0	23,5	18,5	13,2	9,1	6,3	3,2
	,022	,069	,176	,293	,405	,470	,521	,574	,614	,643	,674
G5S4	61,4	57,6	47,8	32,5	19,9	11,0	9,7	6,3	5,0	5,1	3,9
	,025	,056	,153	,306	,432	,522	,534	,569	,581	,581	,593
G5S5	76,8	69,9	56,3	34,6	23,8	15,6	7,5	6,0	4,2	2,9	2,0
	,017	,081	,218	,435	,543	,624	,705	,721	,739	,751	,761
G6S1	86,3	80,3	70,4	49,4	36,2	30,2	22,9	18,8	16,0	13,0	7,7
	,009	,062	,161	,371	,504	,564	,636	,677	,705	,735	,788
G6S2	74,7	65,0	58,7	40,7	32,2	21,6	13,8	9,0	5,7	2,4	1,4
	,022	,113	,176	,356	,440	,546	,624	,673	,705	,739	,749
G6S3	89,6	84,6	72,3	54,3	40,2	28,1	23,2	17,3	12,4	7,5	1,1
	,018	,063	,187	,366	,507	,628	,677	,737	,785	,834	,898
Istotne	0	3	11	16	18	20	20	20	20	20	20

Tabela 5.2. Podsumowanie wyników ANOVA dla obrazów o zredukowanym poziomie szarości

Bit	1	2	3	4	5	6	7	8	9	10	11
G1S1	90,7	91,4	89,9	90,5	83,0	68,3	4,9	0,0	0,0	0,0	0,1
	,014	,009	,017	,015	,081	,228	,862	,911	,911	,911	,910
G1S2	81,7	79,7	78,5	72,7	57,9	45,2	29,0	27,1	24,5	20,1	14,1
	,016	,033	,044	,103	,250	,377	,540	,558	,584	,629	,689
G1S3	91,9	92,2	92,1	94,1	88,0	78,8	40,1	12,7	15,1	9,5	9,3
	,003	,001	,003	,026	,042	,133	,520	,795	,771	,827	,829
G2S1	71,1	70,4	69,1	69,5	65,7	55,6	49,3	34,0	27,2	2,2	0,0
	,009	,010	,022	,026	,054	,154	,218	,370	,438	,688	,710
G2S2	89,7	89,8	89,5	89,7	89,5	88,8	88,0	84,7	75,4	24,8	4,2
	,002	,002	,003	,002	,004	,009	,017	,050	,143	,649	,855
G2S3	79,1	78,9	78,9	78,8	79,0	78,9	79,9	77,8	56,3	27,2	5,7
	,003	,004	,003	,004	,004	,003	,010	,019	,228	,518	,734
G2S4	82,4	81,7	80,1	82,2	79,6	75,7	73,8	71,4	52,0	8,0	0,6
	,008	,010	,026	,009	,035	,069	,089	,112	,306	,746	,820
G3S1	86,7	85,9	87,5	87,1	87,0	85,1	55,1	42,5	39,8	32,7	16,5
	,013	,021	,005	,009	,010	,028	,328	,454	,481	,552	,714
G3S2	92,8	92,8	92,6	92,3	89,7	82,4	59,3	19,9	8,5	7,0	6,0
	,007	,008	,008	,012	,036	,109	,339	,734	,848	,862	,872
G3S3	78,2	79,1	79,1	78,7	76,3	69,7	49,1	25,7	4,6	1,6	1,2
	,005	,006	,006	,002	,030	,089	,296	,529	,741	,770	,776
G3S4	84,8	84,4	85,1	84,0	80,0	73,9	48,5	23,2	19,2	14,3	13,5
	,006	,010	,008	,011	,048	,109	,363	,616	,656	,705	,713
G4S1	91,0	91,2	91,0	90,8	91,0	90,3	86,4	87,4	78,1	37,9	9,7
	,003	,002	,003	,005	,004	,009	,047	,038	,131	,533	,815
G5S1	61,4	62,7	61,5	60,1	53,1	43,9	31,5	20,9	12,0	8,0	6,1
	,011	,012	,011	,021	,089	,181	,305	,411	,500	,540	,559
G5S2	88,3	88,6	88,2	86,2	84,4	76,8	46,0	18,6	7,8	9,0	8,6
	,009	,011	,010	,014	,031	,107	,415	,689	,797	,785	,790
G5S3	70,2	69,4	70,7	68,7	67,6	60,6	41,1	20,2	10,9	7,0	7,1
	,013	,016	,013	,020	,032	,099	,294	,503	,596	,636	,635
G5S4	62,8	64,1	61,8	63,8	60,6	52,1	35,0	13,0	10,2	5,8	2,8
	,010	,016	,020	,023	,029	,111	,282	,501	,529	,573	,603
G5S5	77,5	78,8	77,2	76,8	74,0	66,3	36,3	16,6	7,9	3,9	2,3
	,010	,015	,013	,015	,041	,117	,417	,615	,702	,741	,757
G6S1	86,9	86,5	82,2	76,7	69,3	38,3	19,9	11,9	11,0	10,1	9,7
	,009	,002	,043	,099	,172	,482	,666	,746	,755	,764	,768
G6S2	75,6	75,7	75,4	74,6	70,3	30,6	27,4	24,1	23,2	26,4	27,1
	,006	,006	,010	,019	,063	,456	,488	,521	,530	,498	,492
G6S3	93,4	91,6	90,9	89,0	76,5	60,3	5,3	0,1	0,1	0,1	0,1
	,025	,009	,002	,021	,145	,306	,857	,909	,909	,908	,908
Istotne	0	0	0	0	1	8	16	17	18	20	20

Tabela 5.3. Podsumowanie wyników ANOVA dla obrazów ze sztuczną okluzją

Segment	A1	A2	A3	B1	B2	B3	C1	C2	C3
G1S1	90,4	81,0	92,1	90,7	25,5	90,8	88,5	59,3	78,9
	,011	,101	,012	,012	,656	,018	,026	,318	,122
G1S2	83,2	59,1	78,5	80,9	23,8	78,1	82,7	65,5	80,1
	,004	,238	,046	,021	,592	,052	,005	,174	,029
G1S3	89,9	75,3	88,7	91,8	37,7	86,9	91,7	61,2	78,8
	,024	,170	,035	,005	,545	,053	,006	,311	,134
G2S1	70,5	55,2	69,3	69,2	7,8	70,4	71,2	57,2	69,2
	,007	,158	,024	,022	,632	,015	,004	,140	,020
G2S2	89,9	75,5	89,2	89,7	16,8	90,3	89,6	79,7	90,2
	,004	,142	,005	,001	,729	,007	,001	,100	,006
G2S3	79,5	61,8	78,9	79,0	26,8	79,2	79,1	59,6	77,7
	,010	,173	,006	,002	,523	,005	,002	,195	,013
G2S4	82,8	79,4	82,6	81,6	8,9	82,6	82,7	67,0	74,7
	,005	,035	,015	,012	,737	,016	,009	,156	,079
G3S1	87,8	42,6	86,6	87,4	0,0	88,0	87,6	87,4	86,3
	,003	,453	,013	,006	,879	,004	,004	,007	,018
G3S2	93,3	85,0	92,3	92,6	12,9	92,5	92,7	72,7	92,4
	,001	,083	,010	,007	,804	,009	,006	,206	,009
G3S3	78,6	71,3	76,0	78,9	9,8	77,9	78,3	69,3	79,0
	,001	,077	,034	,004	,688	,015	,004	,093	,007
G3S4	84,7	79,3	84,8	84,2	6,8	84,8	85,0	71,2	84,8
	,005	,055	,014	,011	,780	,015	,007	,136	,010
G4S1	91,2	85,5	91,0	91,0	16,4	91,3	91,1	75,5	91,2
	,001	,057	,003	,002	,748	,003	,003	,157	,003
G5S1	63,0	45,4	58,3	61,9	25,9	60,1	61,9	40,5	58,9
	,014	,165	,037	,011	,360	,022	,012	,215	,031
G5S2	87,7	65,6	84,4	87,4	27,4	86,8	87,1	66,6	86,0
	,006	,219	,032	,007	,601	,011	,005	,209	,016
G5S3	71,0	56,0	67,1	70,7	22,9	69,1	69,6	46,5	69,0
	,010	,145	,034	,010	,477	,017	,011	,240	,018
G5S4	63,5	51,4	60,8	62,4	15,4	60,2	64,7	52,7	60,8
	,009	,117	,030	,015	,478	,031	,017	,104	,025
G5S5	77,8	56,3	74,6	78,4	1,7	76,3	77,8	65,8	73,9
	,006	,218	,035	,010	,593	,019	,009	,122	,042
G6S1	86,3	75,9	86,5	86,9	13,9	86,5	86,7	70,9	87,5
	,008	,107	,010	,005	,726	,008	,004	,156	,010
G6S2	75,8	65,0	73,7	76,2	9,2	75,6	76,1	65,7	75,3
	,005	,113	,025	,001	,670	,009	,001	,106	,011
G6S3	90,9	92,9	92,5	93,4	7,6	92,0	92,7	80,8	88,9
	,001	,020	,016	,025	,833	,020	,019	,101	,029
Istotne	0	6	0	0	20	0	0	7	0

Tabela 5.4. Podsumowanie wyników ANOVA dla filmów z cyklicznym oślepieniem

Cykl	1-1,0,6	4-1,0,6	1-1,0,8	1-1,1,0	6-1,0,6	4-1,0,8	4-1,1,0	6-1,0,8	6-3,0,8	6-1,1,0	6-3,1,0
G1S1	87,7 ,034	81,6 ,095	82,4 ,087	82,5 ,086	75,4 ,156	74,6 ,165	62,4 ,300	61,0 ,286	42,8 ,483	61,8 ,293	40,0 ,510
G1S2	88,6 ,087	73,4 ,096	71,6 ,114	74,1 ,088	74,1 ,088	66,3 ,166	58,5 ,244	55,3 ,277	53,9 ,290	54,9 ,281	22,3 ,607
G1S3	89,0 ,032	80,7 ,114	83,2 ,090	81,1 ,110	65,4 ,268	66,6 ,255	54,5 ,377	51,2 ,410	53,0 ,392	61,1 ,311	52,1 ,400
G2S1	60,7 ,117	52,0 ,190	52,7 ,183	48,6 ,225	49,8 ,212	49,5 ,215	35,5 ,355	35,0 ,360	28,0 ,431	29,3 ,418	21,1 ,500
G2S2	88,6 ,017	85,3 ,044	84,8 ,048	75,0 ,147	83,4 ,063	73,1 ,166	62,6 ,271	65,7 ,240	56,9 ,328	51,4 ,383	34,9 ,548
G2S3	77,6 ,030	76,3 ,0270	76,4 ,0270	72,0 ,070	71,6 ,075	71,7 ,073	72,8 ,072	70,3 ,087	69,7 ,093	65,4 ,137	52,7 ,264
G2S4	79,4 ,032	70,9 ,116	62,5 ,201	47,9 ,347	60,9 ,217	41,5 ,411	32,6 ,5	34,3 ,483	25,1 ,575	24,1 ,585	16,3 ,663
G3S1	82,6 ,053	83,1 ,048	78,6 ,093	77,4 ,105	81,6 ,063	75,7 ,122	61,0 ,269	56,6 ,313	36,2 ,517	30,0 ,579	24,3 ,636
G3S2	86,7 ,065	82,9 ,104	81,8 ,114	77,4 ,126	78,4 ,149	68,5 ,248	62,5 ,307	57,5 ,358	29,9 ,633	34,3 ,590	19,1 ,742
G3S3	75,0 ,040	61,9 ,167	67,5 ,112	65,5 ,132	65,2 ,135	57,6 ,210	50,2 ,284	37,2 ,414	24,6 ,540	16,1 ,626	3,3 ,753
G3S4	78,3 ,065	75,0 ,098	71,8 ,130	72,0 ,128	72,6 ,122	65,6 ,192	54,4 ,303	51,9 ,329	35,5 ,493	32,0 ,528	23,2 ,616
G4S1	89,6 ,017	89,4 ,020	87,5 ,036	84,6 ,066	89,2 ,021	87,3 ,040	82,6 ,085	85,1 ,061	78,2 ,130	75,1 ,160	61,0 ,301
G5S1	56,8 ,051	54,6 ,074	52,7 ,093	52,0 ,100	51,7 ,103	47,2 ,147	37,9 ,241	34,1 ,279	21,4 ,406	20,3 ,417	16,5 ,454
G5S2	78,8 ,087	78,8 ,088	73,4 ,141	71,1 ,164	74,6 ,129	66,1 ,214	45,9 ,416	44,3 ,433	26,6 ,610	24,3 ,632	15,6 ,719
G5S3	61,5 ,090	63,5 ,071	54,8 ,157	55,7 ,149	54,3 ,163	48,5 ,221	37,3 ,332	34,0 ,365	21,7 ,489	21,0 ,496	14,8 ,557
G5S4	55,7 ,074	52,8 ,104	50,5 ,126	49,7 ,134	44,5 ,187	40,3 ,229	33,5 ,297	31,3 ,318	23,0 ,401	24,1 ,390	17,8 ,453
G5S5	71,3 ,067	63,3 ,148	65,4 ,127	64,9 ,132	60,0 ,180	57,8 ,203	51,9 ,262	51,1 ,270	39,3 ,388	43,4 ,347	31,1 ,470
G6S1	79,8 ,067	79,7 ,068	76,0 ,105	73,6 ,129	74,1 ,124	67,4 ,191	67,2 ,193	60,9 ,256	47,3 ,392	43,9 ,426	38,5 ,480
G6S2	71,8 ,044	70,9 ,054	70,3 ,060	69,1 ,071	67,6 ,086	65,0 ,112	68,2 ,080	55,8 ,204	48,8 ,274	57,7 ,186	60,3 ,159
G6S3	83,5 ,079	73,8 ,171	80,1 ,117	78,2 ,127	75,6 ,153	64,0 ,269	72,3 ,186	71,2 ,197	62,1 ,288	63,6 ,274	36,1 ,549
Istotne	2	3	4	7	9	14	18	17	17	18	18

niż statystycznej, to jednak jest odporne na obserwacje, które odstają od modelu teoretycznego, co czyni je dobrym uzupełnieniem analizy wariancji. Szczególnie istotnym podparciem okazuje się w przypadku, gdy nie mamy podstaw do odrzucenia hipotezy zerowej w teście z wysokim ryzykiem błędu typu beta. Metryka Wassersteina pozwala na zmierzenie odległości rozkładów jakości detekcji po wprowadzeniu zaburzeń od rozkładu oryginalnego. W tabelach 5.1, 5.2, 5.3, 5.4 czarnym kolorem wypisano wartości odległości Wassersteina, odpowiadające kolejnym poziomom degradacji jakości dla wszystkich podgrup. Można zauważyć, że wraz ze wzrostem poziomu zakłóceń obrazu empiryczne rozkłady jakości próbek oddalają się od rozkładu oryginalnego, w sensie tej metryki. Ponieważ metryka optymalnego transportu jest funkcją odległości, należy brać pod uwagę, że jej wartość nie niesie informacji o kierunku odchylenia od jakości oryginalnej, a ten musi być odczytany z poprzedniej analizy parametrów lub z dystrybuant rozkładów. Na rysunkach 5.13, 5.18, 5.15 i 5.16 znajdują się wykresy odległości Wassersteina dla wszystkich podgrup (zaznaczonych różnymi kolorami) w przykładowych grupach. Po prawej stronie każdego wykresu można znaleźć dystrybuantę rozkładu jakości dla tej grupy. Można z niej odczytać, jakie jest prawdopodobieństwo rozpoznania obiektu z zadaną jakością lub mniejszą. Ma ona stanowić informację o kierunku w jakim odsuwa się, w sensie tej metryki, rozkład jakości względem oryginału.

W przypadku eksperymentów związanych z bitami i rozmyciem zaobserwowano najczęściej gładki wzrost odległości rozkładu wraz ze wzrostem poziomu zaburzenia. Analizowany zbiór danych zawiera jednak grupy, dla których w sposób naturalny warunki ekspozycji degradują obraz w podobny sposób jak sztucznie wprowadzone zaburzenia. Można zatem oczekiwać, że gdy naturalne i sztuczne zniekształcenia tego samego rodzaju zostaną na siebie nałożone, wyniki będą się wyróżniać na tle innych grup. Pierwszym przykładem jest sztuczne zamazanie obrazu, wprowadzone do scen nagrywanych podczas deszczu, co powoduje pojawienie się kropel wody rozpryskiwanych z nawierzchni drogi i rozmazujących obraz nagrywany przez kamerę. Okazuje się, że ponieważ widok jest już zaburzony z naturalnych powodów, wprowadzenie sztucznego zaburzenia nie wpływa tak mocno na pogorszenie jakości detekcji, jak w przypadku scen, gdzie brak naturalnego rozmycia. Aby to zaobserwować należy porównać rysunek 5.13 dla G3 z wynikami przedstawionymi na rysunku 5.14 dla G1. Dla G3, która reprezentuje typowe wyniki otrzymane w pozostałych grupach, oprócz grupy G1, na większość podgrup rozmycie wpływa już przy wykorzystaniu macierzy o rozmiarze 9 pikseli. Dla grupy G1, czyli grupy scen nagrywanych oryginalnie w warunkach, które mogą powodować powstawanie kropel wody na soczewce kamery, rozmycie nie wpływa istotnie na jakość detekcji aż do

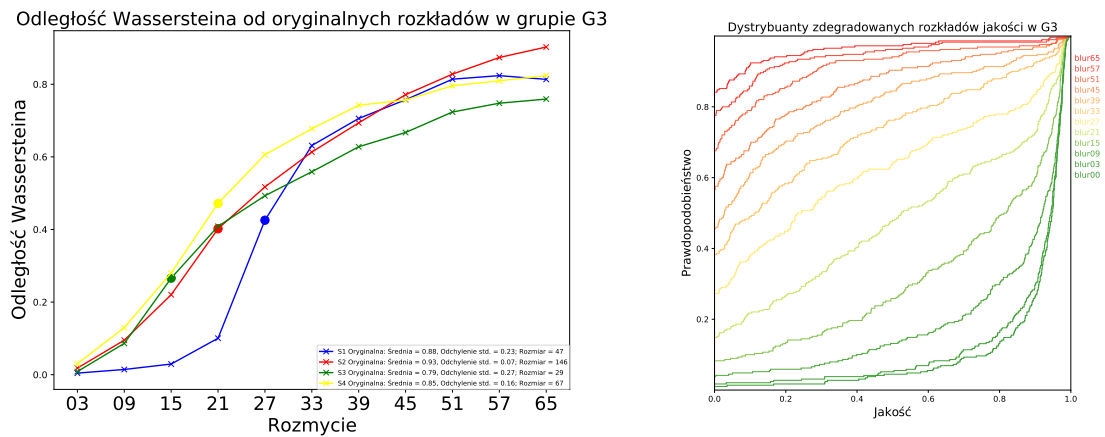
macierzy o rozmiarze 21 pikseli. Jedynym wyjątkiem jest podgrupa G1S2 i jest to spowodowane najprawdopodobniej przez fakt, że jakość oryginalna w scenach tej podgrupy jest znacznie słabsza od pozostałych.

Na rysunku 5.17 zamieszczono wartości metryki Wassersteina obliczone dla wszystkich podgrup danych wykorzystanych w eksperymencie. Analiza wariancji potwierdziła istotność różnicy dla przypadku wyzerowania 6 ostatnich bitów obrazu, co odpowiada wartości 0,159 średniej odległości Wassersteina. Dla przypadku gdzie wyzerowano tylko 5 najmniej znaczących bitów, średnia wartość tej odległości nie przekracza 0,1, jednak tylko dla przypadków zerowania 1, 2, 3 i 4 najmniej istotnych bitów, wszystkie obserwacje są zamknięte w tym promieniu. Na wskazanym wykresie można również zaobserwować odporność na tego typu zaburzenia w grupach G2 oraz G4, której regularność opisano dokładniej powyżej.

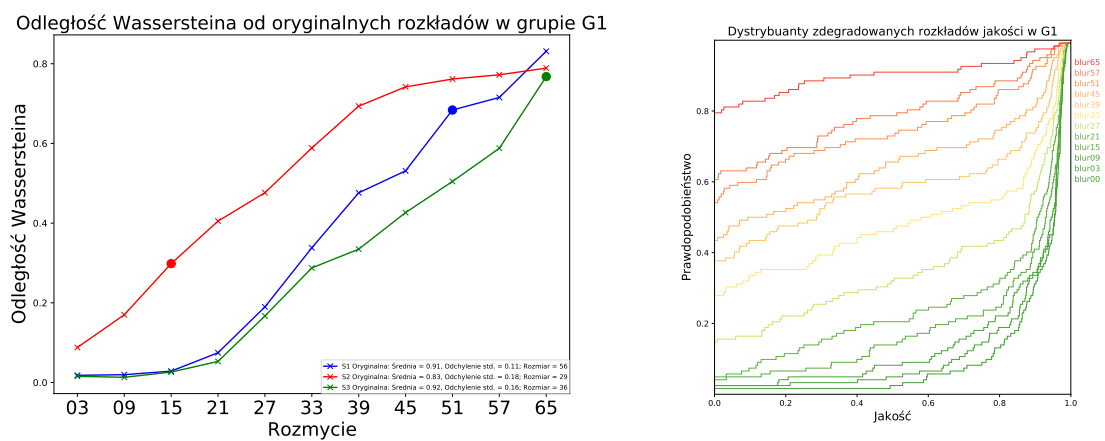
Podobny efekt możemy zaobserwować dla scen nagrywanych w nocy w grupie G2 i scen nagrywanych podczas wschodu lub zachodu słońca w grupie G4, po kolejnych redukcjach głębi bitowej. Spadek jakości w scenach z referencyjnych grup rozpoczyna się w okolicy 5 zredukowanych bitów z gwałtownym spadkiem przy bicie 7 (na przykład rysunek 5.18), gdy dla scen z wyszczególnionych grup spadek zaczyna się nawet dwa bity później, a najgwałtowniejszy spadek przesuwa się na bit 10 (rysunek 5.19).

W przypadku elips możemy zauważyć, że odległość rozkładu jakości od stanu oryginalnego po przesłonięciu centrum obrazka, jest ponad trzykrotnie większe, niż w przypadku segmentów centralnych po lewej i prawej. Dla wszystkich pozostałych segmentów, odległość Wassersteina jest prawie zerowa, co podkreśla wnioski z analizy wariancji.

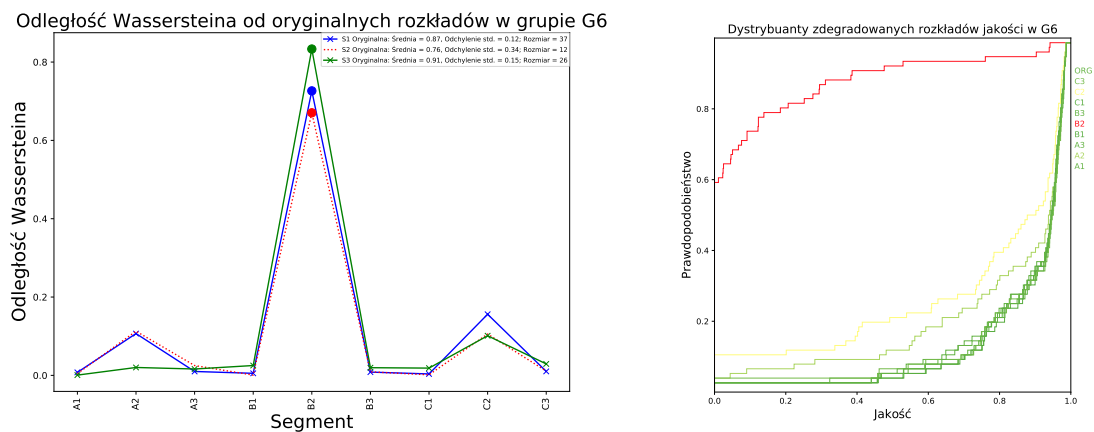
Natomiast wyniki analizy topologicznej danych dla różnych cykli oślepienia pozwalają zauważyć, że największe skoki odległości rozkładu od oryginału otrzymano przy pełnym oślepieniu, ale przedłużanie trwania pośredniej mocy oślepienia ma potencjał bycia porównywalnie niebezpiecznym dla możliwości rozpoznawania obiektów.



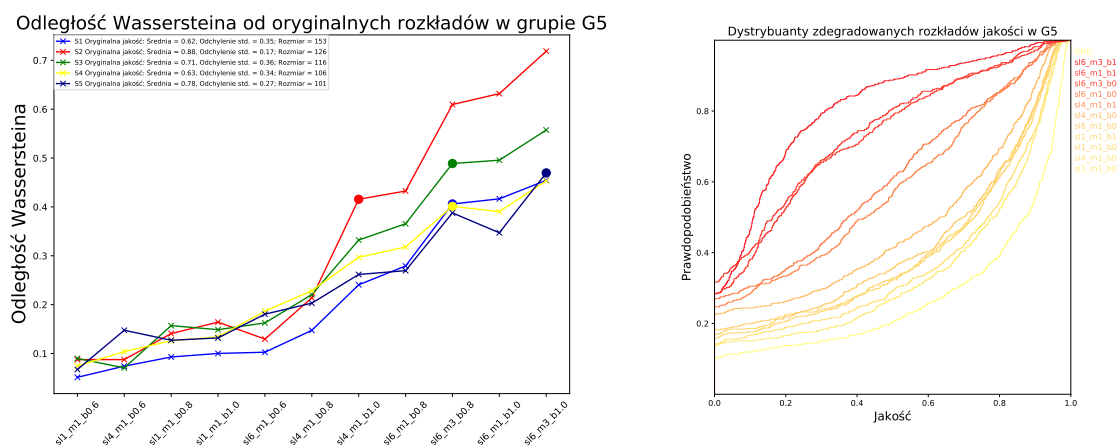
Rysunek 5.13. Przykład wzrostu odległości Wassersteina dla postępującej intensywności rozmycia w grupie G3



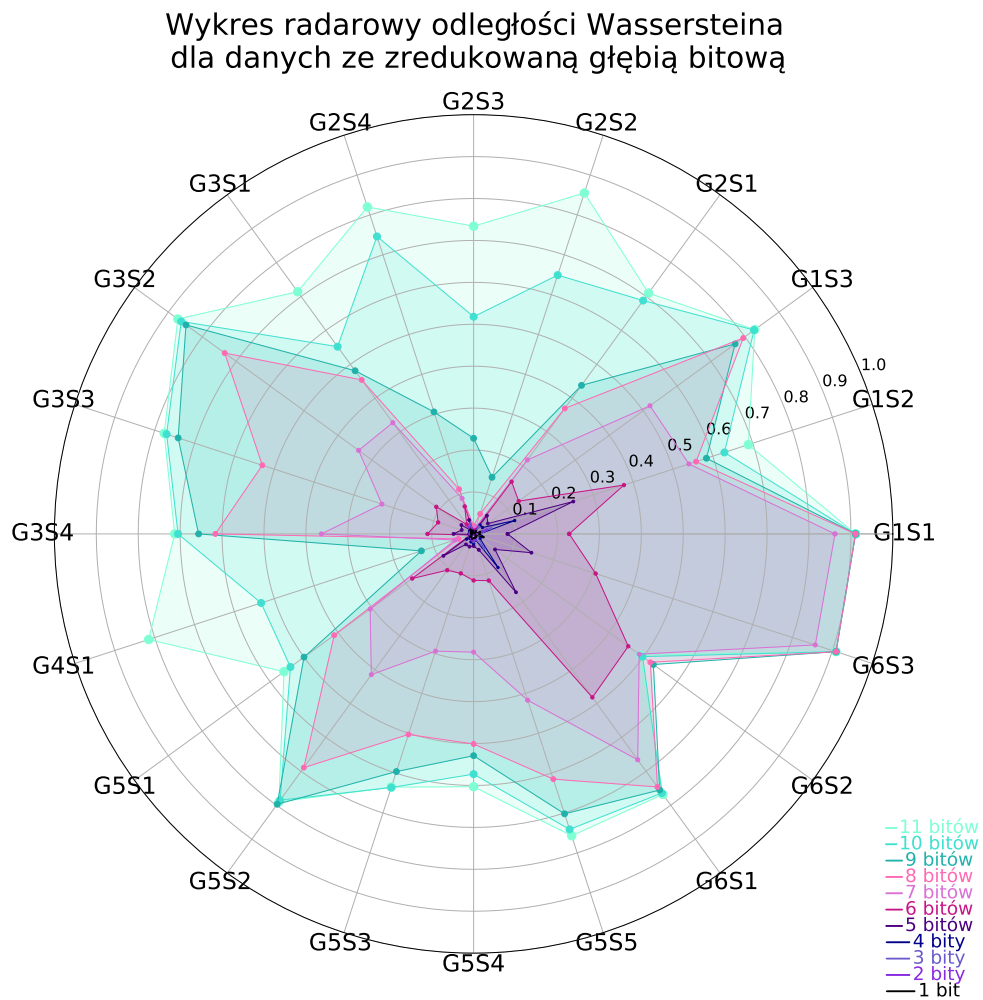
Rysunek 5.14. Wartości Wassersteina dla postępującej intensywności rozmycia w grupie G1. Sztuczna degradacja nakłada się z naturalną



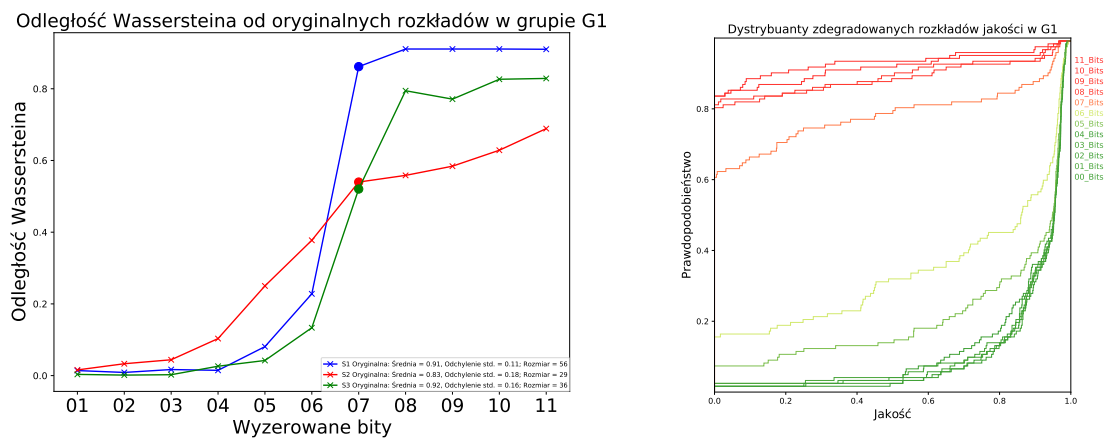
Rysunek 5.15. Przykład wartości Wassersteina dla przesłoniętych segmentów obrazu w grupie G6



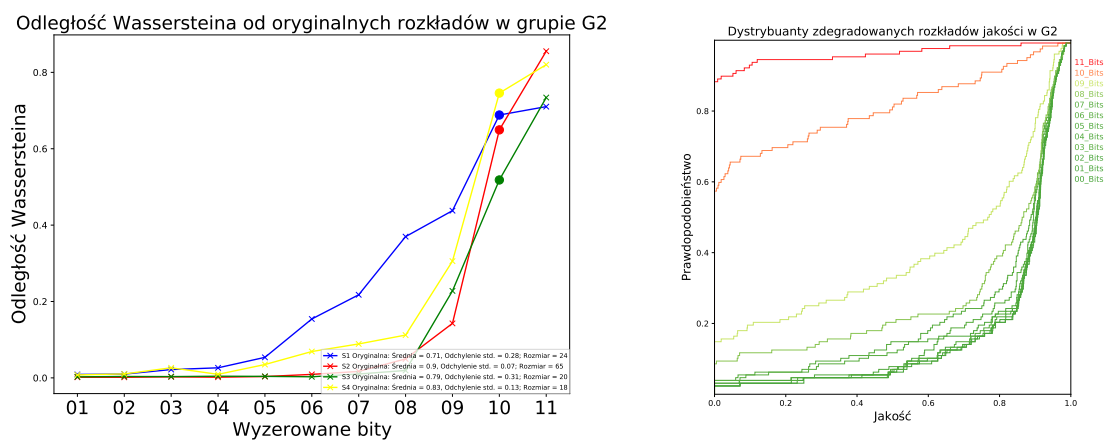
Rysunek 5.16. Przykład wzrostu odległości Wassersteina dla różnych rodzajów cyklicznego osłepienia w grupie G5



Rysunek 5.17. Odległość Wassersteina między oryginalną jakością i jakością uzyskaną na filmach ze zredukowaną głębią bitową we wszystkich podgrupach eksperymentu



Rysunek 5.18. Przykład wzrostu odległości Wassersteina dla postępującej redukcji głębi bitowej w grupie G1



Rysunek 5.19. Wartości Wassersteina dla redukcji bitów w grupie G2. Sztuczna degradacja nakłada się z naturalną

Rozdział 6

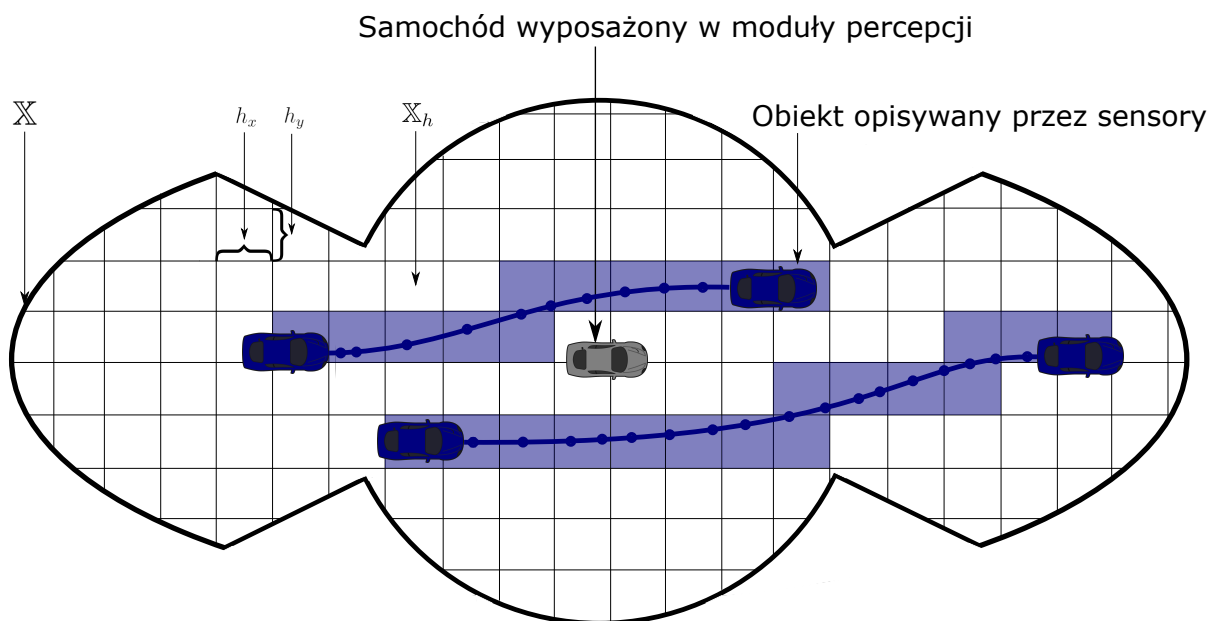
Charakterystyka dużych zbiorów trajektorii

6.1. Rekonstrukcja trajektorii ze zbiorów detekcji

Analiza śledzenia obiektów jest związana z ekstrakcją pełnej trajektorii ruchu obiektu z sekwencji skanów radarowych lub lidarowych dostarczających, w regularnych odstępach czasu, informacji o położeniu obiektu względem czujnika. Taka analiza może również być wykonana w oparciu o dane z kamery, ze względu na możliwość estymacji odległości od kamery rozpoznawanych obiektów. Pojedynczy zestaw informacji zawiera w sobie, oprócz pozycji względnej, również informacje dotyczące kinematyki, geometrii i innych danych o stanie obiektu. W tym miejscu zajmiemy się wyłącznie informacją o dwuwymiarowej pozycji środka obiektu względem samochodu, wyposażonego w czujniki widzianej z perspektywy lotu ptaka. Za trajektorię danego obiektu uznaje się ciąg takich pozycji tworzący ślad obecności obiektu, aktualizowanych w równych odstępach czasu $t' \in \mathbb{R}^+$. Wtedy ciągłą trajektorią ruchu obiektu o numerze j nazywa się funkcję:

$$\mathcal{T}_j : [k_j t', m_j t'] \rightarrow \mathbb{X} \subset \mathbb{R}^2, \quad (6.1)$$

gdzie $k_j, m_j \in \mathbb{N}$, $k_j < m_j$ to początkowe i końcowe numery próbkowania, w których obiekt j był opisywany przez moduły percepcji. Natomiast \mathbb{X} to przestrzeń stanów składająca się z wszystkich pozycji $[x, y]$ na planie dwuwymiarowym, które znajdują się w polu widzenia sensora zlokalizowanego w punkcie $[0, 0]$. Dla dowolnego $n \in \{k_j, \dots, m_j\} \subset \mathbb{N}$ wartość $\mathcal{T}_j^{(s)} = \mathcal{T}_j(nt')$ jest zdefiniowana jako pozycja $[x_j(nt'), y_j(nt')]$ zwrócona w danym momencie nt' przez sensor. Natomiast w pozostałych przypadkach tworzących ciągłą



Rysunek 6.1. Siatka zajętości przetransformowana z przestrzeni stanów \mathbb{X} zawartej w polu widzenia sensorów samochodu autonomicznego stanowiącego punkt odniesienia.

Obiekty poruszające się względem niego podczas scenariusza testowego generują układy trajektorii

trajektorię ruchu:

$$\forall t \in [k_j t', m_j t'] \exists n \in \{k_j, \dots, m_j - 1\} : t \in [n, n + 1], \quad (6.2)$$

a wartość:

$$\mathcal{T}_j(t) = [x_j(t), y_j(t)] \quad (6.3)$$

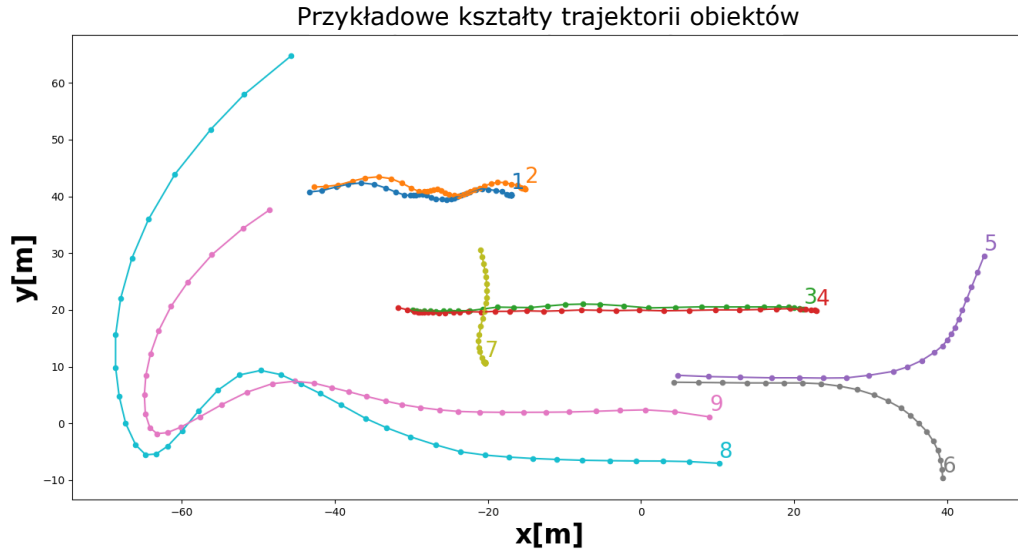
otrzymuje się, przy założeniu:

$$(y_j(t) - y_j(nt'))(x_j((n+1)t') - x_j(nt')) = (y_j((n+1)t') - y_j(nt'))(x_j(t) - x_j(nt')). \quad (6.4)$$

W ten sposób można uzyskać ciągłą trajektorię w czasie t opisaną na pozycjach zwróconych przez sensor.

6.1.1. Siatka zajętości

Aby podjąć analizę trajektorii w otoczeniu samochodu, trajektorie te będą rozpatrywane na uogólnieniu przestrzeni \mathbb{X} na *siatkę zajętości* G_h , złożoną z komórek siatki $G_h(i)$.



Rysunek 6.2. Przykładowe trajektorie zrekonstruowane dla danych radarowych, kropki oznaczają pojedyncze obserwacje pozycji dla obiektów poruszających się w otoczeniu sensora

Definiuje się je jako:

$$G_h(i) = \left\{ [x, y] \in \mathbb{R}^2 : \left\lfloor \frac{x}{h_x} \right\rfloor = i_1, \left\lfloor \frac{y}{h_y} \right\rfloor = i_2 \right\}, \quad (6.5)$$

gdzie $h = [h_x, h_y]$, $h_x > 0$, $h_y > 0$ to rozmiar komórki, $i = [i_1, i_2] \in \mathbb{Z}^2$, a $\lfloor \cdot \rfloor : \mathbb{R} \rightarrow \mathbb{Z}$ to cecha z liczby. Wykorzystując siatkę zajętości G_h można transformować przestrzeń \mathbb{X} otrzymując przestrzeń:

$$\mathbb{X}_h = \{ i \in \mathbb{Z}^2 : \exists_{[x,y] \in \mathbb{X}} : [x, y] \in G_h(i) \}. \quad (6.6)$$

Liczba elementów siatki zależy od parametru h , który można dobrać w zależności od potrzeb lub wymagań. Liczbę komórek wchodzących w interakcję z ciągłą trajektorią ruchu \mathcal{T}_j , można obliczyć jako moc zbioru:

$$V_h(\mathcal{T}_j) = \{ i \in \mathbb{X}_h : \exists_{t \in [k_j t', m_j t']} : [x_j(t), y_j(t)] \in G_h(i) \}. \quad (6.7)$$

6.1.2. Podobieństwo trajektorii

Poniżej przedstawiono przykłady metryk i miar podobieństwa trajektorii pochodzących z pracy [154] i przeglądu z sekcji 3.3.3. Ich wzory jednak dostosowano do porównywa-

nia trajektorii koncentrującego się na analizie zajętości generowanej przez nie na siatce G_h wokół samochodu. Do obliczenia wartości metryk wybrano zestaw trajektorii pochodzących z opublikowanego przez Toyota i Instytut Technologii Karlsruhe zbioru KITTI [155]. Trajektorie mają różną długość, kształt, pozycję względem siebie i mogą należeć do tej samej lub różnych klas wykrywanych obiektów. Zostały przedstawione na rysunku 6.2.

Metryka indukowana przez normę L^p

Pierwsza definicja opiera się na odległości Minkowskiego między dwoma punktami w unormowanej przestrzeni wektorowej z normą L^p :

$$D_{L^p}(\mathcal{T}_i, \mathcal{T}_j) = \left(\sum_{t_i, t_j} (|x_i(t_i) - x_j(t_j)|^p + |y_i(t_i) - y_j(t_j)|^p) \right)^{\frac{1}{p}},$$

gdzie metoda doboru par t_i, t_j może być różna. Bez utraty ogólności można założyć, że $V_h(\mathcal{T}_i) > V_h(\mathcal{T}_j)$. Dla każdego punktu rozpinającego trajektorię dominującą \mathcal{T}_i paruje się najbliższy leżący punkt trajektorii pokrywającej mniejszą powierzchnię siatki zajętości \mathcal{T}_j . Można manipulować parametrem p otrzymując różne własności wzoru. Dla $p = 1$ odległość jest podobna do odległości Manhattańskiej. Dla $p = 2$ jest analogiczna do odległości Euklidesowej, a jeśli $p = \infty$ otrzymuje się metrykę maksimum Czebyszewa. W tabeli 6.1 przedstawiono rezultaty porównania zestawu trajektorii z rysunku 6.2 według tej definicji przy założeniu $p = 2$.

Tabela 6.1. Wartości podobieństwa (%) w sensie normy L^p dla trajektorii przedstawionych na rysunku 6.2

Trajektoria	1	2	3	4	5	6	7	8	9
1	0,0	9,9	165,5	178,9	281,5	282,7	131,8	275,7	251,1
2	9,9	0,0	163,9	177,2	276,3	279,5	136,6	287,9	258,1
3	165,5	163,9	0,0	5,4	150,0	141,3	108,2	211,8	190,2
4	178,9	177,2	5,4	0,0	141,6	133,7	82,9	230,1	205,4
5	281,5	276,3	150,0	141,6	0,0	75,5	257,8	283,8	271,0
6	282,7	279,5	141,3	133,7	75,5	0,0	231,4	234,4	229,4
7	131,8	136,6	108,2	82,9	257,8	231,4	0,0	173,9	153,4
8	275,7	287,9	211,8	230,1	283,8	234,4	173,9	0,0	49,3
9	251,1	258,1	190,2	205,4	271,0	229,4	153,4	49,3	0,0

Kolejny wzór to miara podobieństwa oparta o Indeks Jaccarda. Opisuje ona stosu-

nek fragmentu siatki, pokrytego równocześnie przez obie porównywane trajektorie, do fragmentu zawierającego je obie w całości.

Podobieństwo trajektorii w sensie Jaccarda

$$D_J(\mathcal{T}_i, \mathcal{T}_j) = \frac{|V_h(\mathcal{T}_i) \cap V_h(\mathcal{T}_j)|}{|V_h(\mathcal{T}_i) \cup V_h(\mathcal{T}_j)|},$$

Wartości tej miary obliczone dla przykładowego zestawu trajektorii znajdują się w tabeli 6.2.

Tabela 6.2. Wartości podobieństwa (%) w sensie Jaccarda dla trajektorii przedstawionych na rysunku 6.2

Trajektoria	1	2	3	4	5	6	7	8	9
1	100	66,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0
2	66,7	100	0,0	0,0	0,0	0,0	0,0	0,0	0,0
3	0,0	0,0	100	17,9	0,0	0,0	3,8	0,0	0,0
4	0,0	0,0	17,9	100	0,0	0,0	3,3	0,0	0,0
5	0,0	0,0	0,0	0,0	100	2,9	0,0	0,0	0,0
6	0,0	0,0	0,0	0,0	2,9	100	0,0	0,0	0,0
7	0,0	0,0	3,8	3,3	0,0	0,0	100	0,0	0,0
8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	100	4,1
9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	4,1	100

Odległość trajektorii w sensie Hamminga

Można również zastosować podejście analogiczne jak w przypadku obliczania odległości Hamminga porównującej dwa binarne ciągi danych. Jest ona równa liczbie bitów niezgodnych pomiędzy dwoma ciągami. W naszym przypadku wartością odległości będzie liczba komórek siatki zawierających tylko jedną z dwóch porównywanych trajektorii:

$$D_H(\mathcal{T}_i, \mathcal{T}_j) = |V_h(\mathcal{T}_i) \cup V_h(\mathcal{T}_j)| - |V_h(\mathcal{T}_i) \cap V_h(\mathcal{T}_j)| \quad (6.8)$$

Wartości tej miary obliczone dla przykładowego zestawu trajektorii znajdują się w tabeli 6.3.

Tabela 6.3. Wartości podobieństwa (%) w sensie miary Hamminga dla trajektorii przedstawionych na rysunku 6.2

Trajektoria	1	2	3	4	5	6	7	8	9
1	0	6	36	40	33	33	21	54	52
2	6	0	36	40	33	33	21	54	52
3	36	36	0	32	39	39	25	60	58
4	40	40	32	0	43	43	29	64	62
5	33	33	39	43	0	34	24	57	55
6	33	33	39	43	34	0	24	57	55
7	21	21	25	29	24	24	0	45	43
8	54	54	60	64	57	57	45	0	70
9	52	52	58	62	55	55	43	70	0

6.2. Analiza scenariuszy testowych

Scenariusz testowy \mathcal{S} w niniejszym rozdziale rozumiany jest jako zbiór trajektorii ruchu, pobranych podczas spójnej sceny testowej:

$$\mathcal{S} = \{T_j : j \in J \subset \mathbb{N}_+\}. \quad (6.9)$$

Dla każdego scenariusza testowego można wyznaczyć jaką część siatki zajętości jest przez trajektorie należące do niego okupowana według wzoru:

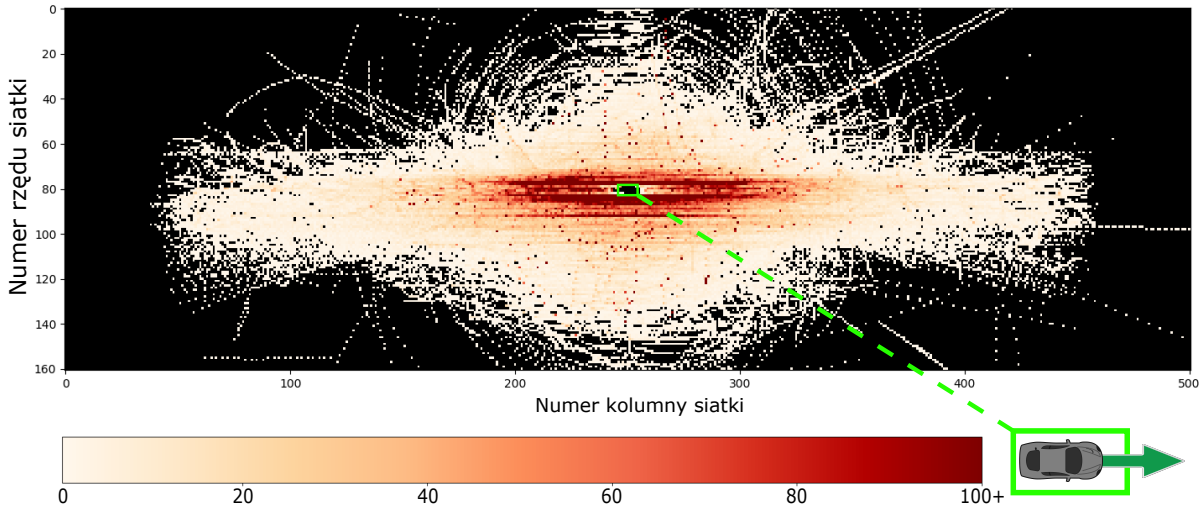
$$C_h(\mathcal{S}) = \frac{\left| \bigcup_{j \in J} V_h(T_j) \right|}{|\mathbb{X}_h|}. \quad (6.10)$$

Wyznacznik ten nie niesie jednak informacji o intensywności okupowania danego miejsca na siatce. Dlatego nie można na jego podstawie rozróżnić jak wiele obiektów lub przez jak długi czas znajdowało się w danej strefie wokół samochodu.

6.2.1. Mapa termiczna zajętości otoczenia

Dlatego do dalszej analizy wykorzystano pojęcie *histogramu zajętości* definiowanego jako funkcję $H : \mathbb{X}_h \rightarrow \mathbb{N}$, gdzie

$$H(i) = \sum_{T_j \in \mathcal{S}} |\{t \in \{k_j t', \dots, m_j t'\} : T_j(t) \in G_h(i)\}|. \quad (6.11)$$



Rysunek 6.3. Mapa termiczna zajętości otoczenia wokół samochodu znajdującego się w centrum układu współrzędnych i skierowanego w prawą stronę wykresu. Wykreślona ona została dla wszystkich trajektorii ze zbioru scenariuszy testowych

Takie podejście otwiera nowe możliwości wizualizacji scenariusza testowego oraz zbioru scenariuszy. Najskuteczniejszą metodą wizualizacji histogramów dwuwymiarowych jest mapa termiczna. Jest to technika wizualizacji danych, która pokazuje natężenie zjawiska na dwuwymiarowej przestrzeni za pomocą intensywności koloru, która zmienia się wraz ze wzrostem wartości funkcji dla wektora dwuwymiarowego. Na rysunku 6.3 przedstawiono przykład mapy termicznej zajętości dla zbioru wszystkich scenariuszy testowych pochodzących z niepublikowanej bazy danych Aptiv, wykorzystując $h_x = 1m, h_y = 1m$. Przyjęto następującą konwencję: kolor czarny oznacza, że scenariusz (lub suma zbioru scenariuszy w tym konkretnym przypadku) nie zawiera trajektorii wchodzących w interakcję z daną komórką siatki. Komórki siatki, które przecinają się wyłącznie raz z jedną trajektorią, są białe. Kolor biały stopniowo zmienia się na czerwony dla komórek siatki, które w sumie były zajęte przez większą liczbę trajektorii. Wartość graniczna dla tej zmiany wynosi 100, jak pokazano na legendzie koloru.

W oparciu o histogram zajętości dla dowolnego scenariusza testowego \mathcal{S} , definiuje się rozkład zajętości $\mu : \mathbb{X}_h \rightarrow [0, 1]$ w następujący sposób

$$\mu(i) = \frac{H(i)}{\sum_{j \in \mathbb{X}_h} H(j)}, i \in \mathbb{X}_h. \quad (6.12)$$

Warto zwrócić uwagę, że rozkład zajętości jest gęstością dwuwymiarowego dyskretnego

rozkładu prawdopodobieństwa, natomiast (\mathbb{X}_h, d) jest przestrzenią metryczną, gdzie d jest indukowaną odległością Euklidesową.

6.2.2. Porównywanie scenariuszy testowych

Zastosowanie rozkładu zajętości scenariuszy umożliwia również ich formalne porównywanie. Odległość między scenariuszami definiuje się jako odległość Wassersteina pomiędzy odpowiadającymi im rozkładami zajętości μ i ν . Obliczanie jej przebiega podobnie jak to przedstawiono w rozdziale poprzednim, jednak uprzednio należy zauważyć, że:

$$\forall h \in \mathbb{R}_+^2 \exists \eta : \mathbb{M} \rightarrow \mathbb{X}_h, \quad (6.13)$$

gdzie $\mathbb{M} = \{1, \dots, |\mathbb{X}_h|\}$ i η jest bijekcją. Aby wyznaczyć odległość $W(\mu, \nu)$ należy znaleźć optymalną $f^* : \mathbb{M} \times \mathbb{M} \rightarrow [0, 1]$, minimalizującą sumę:

$$\sum_{k=1}^{|\mathbb{X}_h|} \sum_{j=1}^{|\mathbb{X}_h|} f(k, j) d(k, j), \quad (6.14)$$

gdzie $d(k, j)$ jest kwadratem odległości Euklidesowej pomiędzy wektorami $\eta(k)$ i $\eta(j)$. Kryteria dla funkcji f są następujące:

$$\forall k, j \in \mathbb{M} : 0 \leq f(k, j) \leq 1, \quad (6.15)$$

$$\forall k \in \mathbb{M} : \sum_{j=1}^{|\mathbb{X}_h|} f(k, j) = \mu(\eta(k)), \quad (6.16)$$

$$\forall j \in \mathbb{M} : \sum_{k=1}^{|\mathbb{X}_h|} f(k, j) = \nu(\eta(j)), \quad (6.17)$$

$$\sum_{k=1}^{|\mathbb{X}_h|} \sum_{j=1}^{|\mathbb{X}_h|} f(k, j) = 1. \quad (6.18)$$

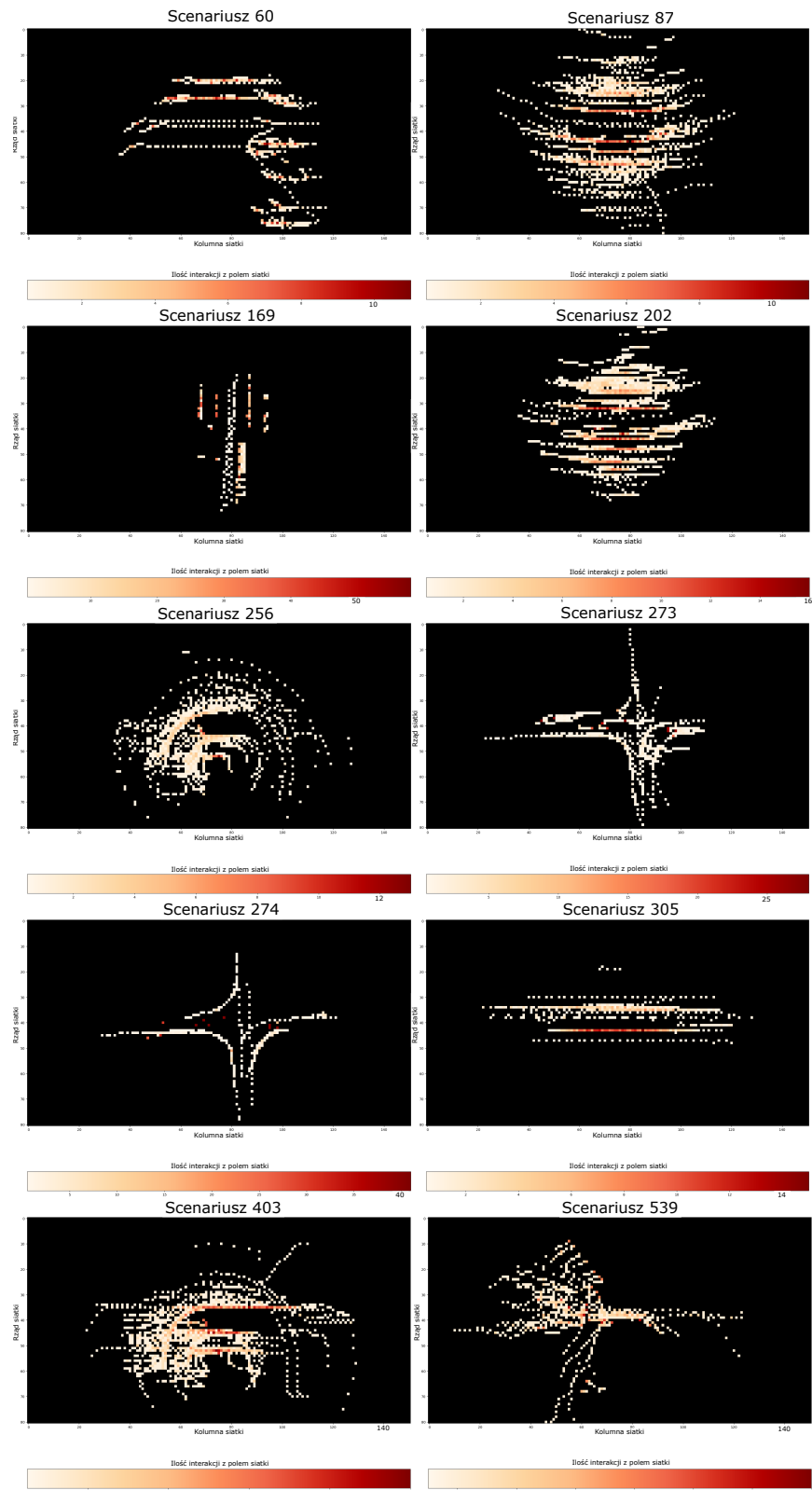
Rozwiązanie takiego układu ponownie wymaga skorzystania z metody Sympleks znajdującej się w [144] i [145]. Wtedy wartość odległości Wassersteina jest wyznaczana ze wzoru:

$$W(\mu, \nu) = \sqrt{\sum_{k=1}^{|\mathbb{X}_h|} \sum_{j=1}^{|\mathbb{X}_h|} f^*(k, j) d(k, j)}. \quad (6.19)$$

Tabela 6.4. Odległości Wassersteina pomiędzy mapami zajętości scenariuszy ze zbioru danych nuScenes

Scena	60	87	169	202	256	273	274	305	403	539
60	0	15,3	14,0	16,5	21,1	14,1	15,6	18,2	20,0	28,7
87	15,3	0	11,2	5,4	9,3	10,2	11,7	11,5	9,5	17,4
169	14,0	11,2	0	10,5	15,7	12,2	13,7	15,6	16,3	24,8
202	16,5	5,4	10,5	0	11,1	11,5	12,9	11,3	11,9	17,5
256	21,1	9,3	15,7	11,1	0	12,0	12,8	12,1	4,9	14,1
273	14,1	10,2	12,2	11,5	12,0	0	4,9	8,8	10,7	20,3
274	15,6	11,7	13,7	12,9	12,8	4,9	0	8,3	10,8	21,4
305	18,2	11,5	15,6	11,3	12,1	8,8	8,3	0	10,4	19,0
403	20,0	9,5	16,3	11,9	4,9	10,7	10,8	10,4	0	15,9
539	28,7	17,4	24,8	17,5	14,1	20,3	21,4	19,0	15,9	0

Taka definicja pozwala - w sposób automatyczny - analizować podobieństwo scenariuszy testowych w oparciu o trajektorie w nim zapisane. Metryka Wassersteina będzie tym większa, im bardziej zróżnicowane są lokalizacje trajektorii znajdujących się w porównywanych zbiorach, po rzutowaniu na siatkę zajętości oraz im bardziej różni się intensywność okupowania przez nie poszczególnych komórek siatki. Aby przetestować tą hipotezę porównano tą metodą zestaw scenariuszy testowych, a wartości zamieszczono w tabeli 6.4. Następnie porównano otrzymane wyniki z wyglądem map termicznych zajętości dla tych scenariuszy 6.4. Zestaw danych, który został wykorzystany do przetestowania metodologii porównywania scenariuszy, to zestaw danych nuScenes [156] dostarczony przez Motional (wspólne przedsięwzięcie firm Hyundai i Aptiv). Pełny zestaw danych zawiera około 1,4 mln obrazów z kamer, 390 tys. skanów lidarowych, 1,4 mln skanów radarowych i 1,4 mln prostokątnych regionów rozpoznania, opisujących obiekty. Zestaw ten składa się na 850 różnych scenariuszy testowych. Największą wartość odległości między scenariuszami, czyli 28,7 uzyskano dla porównania scenariusza 60 i 539. W scenariuszu 60, większość pojazdów poruszała się we frontowej przestrzeni samochodu testowego, podczas gdy w scenariuszu 539 było na odwrót. W scenariuszach 87 i 202, intensywność zajmowania przestrzeni samochodu testowego jest równomiernie rozłożona wokół pojazdu. Wartość odległości między tymi scenariuszami jest względnie niska i wynosi 5,4. Mapy termiczne zajętości z numerami 273 i 274 bazują na danych zebranych na skrzyżowaniach typu X. Odległość Wassersteina obliczona między nimi wynosi 4,9 i wskazuje na ich podobieństwo. Natomiast oba powyższe przykłady scenariuszy na skrzyżowaniach są prawie dwa razy bardziej odległe, w sensie tej metryki, od scenariusza numer 305 nagranych na prostej drodze. Scenariusz ten jest również znacząco odległy (15,6) od tego, gdzie trajektorie

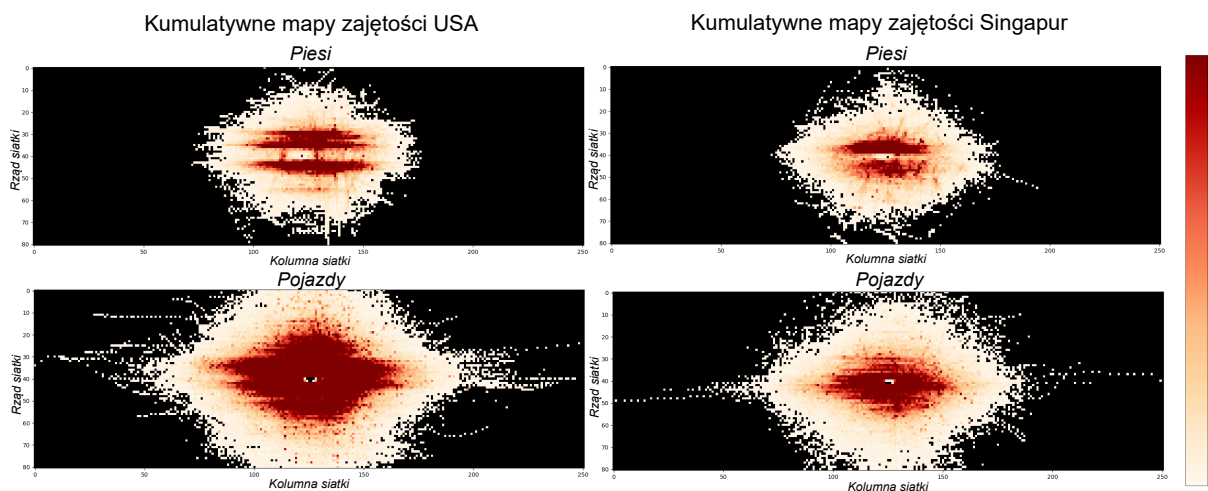


Rysunek 6.4. Przykłady map termicznych zajętości bazujących na scenariuszach z bazy danych nuScenes. Odległości Wassersteina między tymi scenami są zaprezentowane w tabeli 6.4

również poruszały się w linii prostej, lecz w kierunku prostopadłym do kierunku jazdy samochodu testowego (scenariusz 169). Scenariusze, w których samochód testowy wykonuje stanowczy skręt w sytuacji, gdy pozostałe obiekty na drodze kontynuują podróż w stałym kierunku, tworzą charakterystyczne mapy termiczne podobne, do tych z numerami 256 i 403. Taka sytuacja zdarza się na przykład podczas zjazdu z bardzo ruchliwej drogi. Odległość Wassersteina również rozpoznaje wskazane scenariusze jako sytuacje podobne przypisując porównaniu wartość 4,9. Bazując na tym eksperymencie można stwierdzić, że wykorzystana miara odległości skutecznie opisuje podobieństwo między scenariuszami testowymi w oparciu o generowaną przez należące do nich trajektorie intensywność zajętości, natomiast forma reprezentacyjnych map termicznych pozwala analizować szerszy kontekst sytuacji na drodze w jakiej dany scenariusz i jemu podobne był nagrywany.

6.3. Porównanie zbiorów scenariuszy

Typowy projekt dotyczący funkcji wspomagających kierowcę samochodu wymaga zebrania danych pochodzących z setek i tysięcy kilometrów jazd testowych. Samo zapisanie, przechowywanie i wykorzystanie logów tego typu jest sporym wyzwaniem ze względu na dużą ich ilość i złożoną strukturę danych. Często pojemność miejsca dyskowego potrzebnego do przechowania danych projektu wynosi dziesiątki petabajtów. Ze względu na regulacje prawne możliwość przeprowadzenia ponownej weryfikacji i walidacji systemu musi być dostępna wiele lat po zakończeniu projektu, co powoduje istotny wzrost jego kosztów [1]. Dlatego niezwykle istotnym jest by dane, które zebrano i zapisano były wartościowe i wносиły unikalne informacje niezbędne dla rozwoju i testowania modułów percepcji. W wielu przypadkach zbiory scenariuszy jazd testowych pozostają utajnione dla środowisk nie związanych z projektem. Nie mniej niektóre uniwersytety i producenci samochodów decydują się na upublicznienie swoich zestawów, aby mogły być wykorzystane przez komercyjnych badaczy do niezależnego rozwoju sieci neuronowych, technologii i metodologii związanych wykorzystaniem tego typu danych. Przykładami są: zbiór KITTI [155] wydany przez Toyotę i Instytut technologii Karlsruhe, zbiór upubliczniony przez Audi [157], Zbiór Waymo [158], Berkeley DeepDrive (BDD100k) [159], Baidu Apolloscapes [160] i Cityscape Dataset [161]. Zbiór scenariuszy nuScenes [156] wybrano do testowania metodologii w niniejszej pracy, ponieważ zawiera dużą liczbę scenariuszy z anotacjami, umożliwiając ekstrakcję trajektorii różnych typów obiektów. Ponadto można go podzielić w naturalny sposób na dwa podzbiory scenariuszy pochodzące z dwóch różnych krajów, tj. USA i Singapuru, gdzie obowiązują odpowiednio ruch prawostronny



Rysunek 6.5. Kumulatywne mapy termiczne zajętości otoczenia dla trajektorii pojazdów i pieszych zawartych w zbiorach nuScenes zebranych w krajach USA i Singapur

i lewostronny. W niniejszej części pracy przedstawiono metody porównywania dwóch baz danych \mathcal{S} i \mathcal{R} zawierających dane o trajektoriach rozumianych jako zbiory scenariuszy:

$$\mathcal{S} = \{S_1, \dots, S_n\}, \quad (6.20)$$

$$\mathcal{R} = \{R_1, \dots, R_m\}. \quad (6.21)$$

6.3.1. Kumulatywne mapy termiczne

Jako pierwszy sposób porównania zestawów scenariuszy nasuwa się wizualne zestawienie map termicznych zajętości dla trajektorii z nich pochodzących. Aby to zrobić wykreślono mapy nie dla pojedynczych scenariuszy jak na rysunku 6.4, ale zrobiono to na podstawie zbioru trajektorii stanowiącego sumę trajektorii z danego zbioru scenariuszy, czyli jak na rysunku 6.3. Taki sposób wizualizacji nazwano *kumulatywną mapą termiczną zajętości*. W taki sposób porównano zbiory scenariuszy z zestawu nuScenes pochodzące z kraju USA zawierającego 467 scenariuszy i Singapur zawierającego 383 scenariusze. Ponadto uwzględniono podział na trajektorie pieszych i pojazdów. Wyniki zamieszczono na rysunku 6.5. Policzono również wartość metryki Wassersteina dla odpowiadających mapom rozkładów zajętości. Odległość między mapami termicznymi pieszych wyniosła 3,1 natomiast w przypadku pojazdów wyniosła 4,08.

6.3.2. Klastrowanie rozkładów zajętości

Drugim i kluczowym elementem w analizie zbiorów scenariuszy zaproponowanej w tej pracy jest podejście formalne oparte na metryce Wassersteina obliczonej pomiędzy poszczególnymi scenariuszami wchodzącymi w skład zbioru danych. Takie podejście może zapewnić bardziej szczegółowy opis relacji między dwoma dużymi zbiorami danych dotyczących trajektorii. Ideą jest zbadanie, jak zmienia się struktura odległości między scenariuszami po dodaniu do jednego zestawu danych z nowego, złożonego z innych scenariuszy. Aby to zrobić, należy przeprowadzić klastrowanie scenariuszy z zadanego zbioru danych \mathcal{S} , gdzie każdy scenariusz reprezentowany jest przez swój rozkład zajętości, a odległości między nimi są odległościami Wassersteina obliczanymi jak opisano w 6.2.2. Następnie, aby uchwycić wpływ jaki ma na dany zbiór scenariuszy inny zestaw scenariuszy \mathcal{R} , proces klastrowania zostaje powtórzony na sumie dwóch zbiorów scenariuszy $\mathcal{S} \cup \mathcal{R}$. Celem takiego procesu jest sprawdzenie czy \mathcal{R} wnosi do \mathcal{S} nową informację. Innymi słowy czy istotnie go rozszerza pod kątem różnorodności scenariuszy testowych, opisywanych rozkładami zajętości. Ponadto poprzez porównanie dwóch procesów klastrowania oraz ich wyników, możliwym jest formalne opisanie takiego zestawienia dwóch zbiorów scenariuszy. Sam proces klastrowania przebiega dokładnie tak jak algorytm K-średnich, według konwencji opisanych w sekcji 5.1.2, jednak zamiast klastrować jednowymiarowe rozkłady empiryczne jakości, klastrowane są dwuwymiarowe rozkłady zajętości. Podczas przeprowadzania eksperymentu opisanego w następnej sekcji stwierdzono również, że skuteczniejszym wskaźnikiem optymalnej liczby grup w przypadku klastrowania rozkładów zajętości będzie Indeks Dunna [162]. Załóżmy, że dokonano klastrowania zbioru scenariuszy \mathcal{S} . Jako wynik otrzymano klastry scenariuszy $C_1, C_2, \dots, C_e, e \in \mathbb{N}$ oraz odpowiadające im centroidy $\xi_1, \xi_2, \dots, \xi_e$. Każdy centroid $\xi_i, i \in \{1, \dots, e\}$ jest rozkładem prawdopodobieństwa na przestrzeni \mathbb{X}_h , który uśrednia wszystkie rozkłady zajętości z klastra C_i w sensie metryki Wassersteina. Następnie podejmuje się klastrowanie sumy zbiorów scenariuszy z dwóch baz danych $\mathcal{S} \cup \mathcal{R}$. Jako wynik otrzymuje się klastry scenariuszy D_1, D_2, \dots, D_f i odpowiadające im centroidy $\pi_1, \pi_2, \dots, \pi_f$. Różnica między tymi dwoma wynikami algorytmu klastrowania jest informacją jaką baza scenariuszy \mathcal{R} wnosi do bazy \mathcal{S} . Aby formalnie ją opisać, należy zdefiniować przyporządkowanie między zbiorem klastrów uzyskanych z pierwszego przebiegu algorytmu, a zbiorem uzyskanym z drugiego algorytmu. Niech $I_e = \{1, \dots, e\}$ i $I_f = \{1, \dots, f\}$. Definiuje się funkcję $F : I_e \rightarrow I_f$ taką, że:

$$F(i) = \arg \min_{k \in J} W(\xi_i, \pi_k) \quad (6.22)$$

gdzie $J = \{j \in I_f : C_i \cap D_j \neq \emptyset\}$. Oznacza to, że $F(I_e) \subset I_f$, ale niekoniecznie zachodzi równość tych dwóch zbiorów (F nie musi być suriekcją).

6.3.3. Współczynniki różnorodności

Bazując na wynikach klastrowania opisanego powyżej, definiuje się następujące współczynniki różnorodności zbioru scenariuszy testowych \mathcal{R} względem zbioru \mathcal{S} .

Współczynnik wzbogacenia istniejących klastrów:

$$w_1 = \frac{\sum_{i=1}^e W(\xi_i, \pi_{F(i)})}{e}. \quad (6.23)$$

Współczynniki wzbogacenia zbioru klastrów:

$$w_2 = \frac{|f - e|}{e}, \quad (6.24)$$

$$w_2' = \frac{\sum_{i \in I_f - I_e} |D_i|}{|f - e|}, \quad (6.25)$$

$$w_2'' = \frac{\sum_{i \in I_f - I_e} \min_{j \in I_e} W(\xi_j, \pi_i)}{|f - e|}. \quad (6.26)$$

Średnia zmiana inercji wewnętrznej:

$$w_3 = \left| \frac{\sum_{i \in I_e} \sum_{\sigma \in C_i} W(\xi_i, \sigma)^2}{|\mathcal{S}|} - \frac{\sum_{i \in I_e} \sum_{\sigma \in C_i \cup D_{F(i)}} W(\pi_{F(i)}, \sigma)^2}{|\mathcal{S} \cup \mathcal{R}|} \right|. \quad (6.27)$$

W oparciu o powyższe współczynniki zaproponowano metodę porównania dwóch zbiorów scenariuszy. Do eksperymentu wykorzystano ponownie zbiór danych nuScenes. Wybrano z niego 150 scenariuszy pochodzących z USA i 50 zebranych w Singapurze. Celem poniższego eksperymentu jest określenie, na podstawie współczynników różnorodności, czy i w jaki sposób scenariusze zebrane w Singapurze rozszerzają istotnie strukturę scenariuszy jaką tworzy zbiór z USA. Kluczowe statystyki dotyczące wyników klastrowania scenariuszy z USA i sumy scenariuszy z USA i Singapuru znajdują się w tabeli 6.5. Znajdują się w niej również obliczone na podstawie przeprowadzonych procesów wyznaczniki różnorodności. Przed dodaniem zbioru z Singapuru, 150 scenariuszy z USA tworzy 3 kla-

Tabela 6.5. Współczynniki różnorodności wyznaczone dla fragmentów zbioru scenariuszy nuScenes pochodzących z różnych krajów

Zbiory	USA	USA + Singapur
Scenariusze	150	150 + 50
Klastry	3	4
Inercja końcowa	10932, 19	14673, 15
w_1	4, 699	
w_2	0, 333	
w'_2	76, 0	
w''_2	8, 013	
w_3	10921, 349	

stry. Drugi wynik klastrowania składa się z jednej nowej grupy ale zwiększa inercję całej struktury o 34,2%. Średnie przesunięcie centroidów z pozycji jaką mają po pierwszym klastrowaniu (w_1) wynosi 4,7, co oznacza, że zmiana centroidów plasuje się na poziomie odległości między scenariuszami tego samego typu (tabela 6.4) i może być interpretowana jako niska. Jednakże to, że średni rozmiar nowych klastrów czyli (w'_2) jest większy niż zbiór scenariuszy pochodzących z Singapuru wskazuje, że nowe scenariusze rozszerzają pewien zbiór scen ze zbioru USA, który był niewystarczająco reprezentowany by wytworzyć odseparowany klaster w pierwszym procesie. Niska inercja wewnętrzna (w_3) względem inercji końcowej drugiego procesu sugeruje, że nowy klaster jest wciąż mocno niestabilny i ma potencjał do kolejnych rozpadów. Warto zatem rozważyć rozszerzanie zbioru z USA o kolejne próbki z Singapuru.

Rozdział 7

Podsumowanie

Powyższa praca prezentuje propozycje nowych metodologii testowania i ewaluacji danych związanych z działaniem modułów percepcji skupiając się na szerokiej problematyce tych aktywności wynikającej z typów danych, ich różnorodności, ilości i potrzebie wielopoziomowej analizy. Motywacją była chęć ujednolicenia, zagwarantowanie odtwarzalności i poprawa skuteczności ewaluacji modułów percepcji oraz rozbudowanie możliwości wzbogacania puli informacji użytecznych zarówno z punktu widzenia wdrożeniowego jak i badawczego w ramach procesu rozwoju autonomicznych samochodów. Drogą do osiągnięcia tego celu było zaproponowanie formalnych definicji oraz algorytmów wykorzystujących szeroki wachlarz dostępnych narzędzi matematycznych, a także skupienie się na konieczności głębokiego zrozumienia danych w celu ich skutecznej analizy.

7.1. Konkluzja o zrealizowaniu celu pracy

Poniżej znajduje się przegląd osiągnięć i działań podjętych przez Autora w ramach projektu doktorskiego, których podsumowaniem jest niniejsza praca. Autor przedstawił w niej dowody na potwierdzenie tez postawionych w rozdziale pierwszym.

1. W rozdziale czwartym zaproponowano miarę podobieństwa potrzebną w procesie ewaluacji modułów percepcji na tle danych referencyjnych. Jest ona złożona z trzech zaproponowanych miar pomocniczych połączonych z wykorzystaniem ważonej średniej harmonicznej. Jej definicja pozwala na w pełni automatyczną analizę danych w postaci prostokątnych regionów rozpoznania. Zaproponowana konstrukcja miary pozwala dostarczać szczegółowych informacji o stopniu trafności decyzji podejmowanych przez testowany moduł percepcji, co usprawnia proces dopasowania danych referencyjnych i wyników działania sieci przed ostateczną jej ewaluacją. Proces ten

został opisany w przedstawionym algorytmie dopasowania danych. Ponadto możliwość kalibracji pozwala ustalić wpływ poszczególnych aspektów podobieństwa, ocenianych osobno, na finałowy wynik, co pozwala na bardziej precyzyjną definicję akceptowalnych wyników oraz możliwość dopasowania działania ewaluacji do charakterystycznych problemów występujących w różnych modułach percepcji. Zostało to przedstawione na przykładach modułów PED, MOD, TSR i TLR. Ważną zaletą zaproponowanej generalnej miary jakości jest to, że łączy ona podejście holistyczne z możliwością dekompozycji informacji na dobrze izolowane części składowe. Zadbanie o wynik w formacie procentowym, wykorzystującym pełne spektrum przedziału $[0, 1]$, pozwala jasno komunikować rezultaty i ułatwia, następujące w kolejnych procesach, analizę i syntezę wniosków.

Kolejnym problemem, jaki rozwiązano w rozdziale 4. jest analiza jakości sekwencji prostokątów opisujących detekcję obiektów w całym czasie trwania scenariusza testowego. Aby zapobiec degradacji znaczenia kluczowych pierwszych momentów pojawienia się obiektu w zasięgu percepcji pojazdu, zaproponowano miarę podobieństwa podsumowującą lokalne wyniki miar jakości w klatce. Zrobiono to proponując algorytm doboru wag do średniej arytmetycznej w sposób kładący nacisk na relację między pierwszą detekcją, a maksymalnym dopuszczalnym czasem reakcji algorytmów detekcji. W ten sposób otrzymano narzędzie uodporniające algorytm ewaluacji na długotrwały napływ wysokiej jakości danych o drugorzędym znaczeniu.

Zaproponowano też metodę wizualizacji jakości scenariusza testowego umożliwiającą szybką lokalizację problemów z działaniem modułu percepcji, zapewniającą wgląd w kluczowe informacje o sytuacji w jakiej zawiodły, bez potrzeby manualnego przeglądania filmów. Dodatkowo, ponownie wykorzystując miary podobieństwa podjęto się wstępnej analizy rozpoznania fałszywie pozytywnych, które stanowią szczególnie problem z punktu widzenia automatyzacji procesu ewaluacji.

Ostatecznie podjęto się porównania działania zaproponowanych rozwiązań z innymi znanymi podejściami do tych problemów. Szczególną uwagę poświęcono porównaniu z metodami wykorzystującymi indeks Jaccarda. Wskazano przypadki w jakich konieczna jest bardziej szczegółowa analiza by uzyskać poprawną ocenę, która w każdym z tych przypadków jest zapewniana przez zaproponowane podejście. Wyszczególniono jakie własności miar są pożądane, zarówno w przypadku analizy w klatkach jak i w przypadku analizy sekwencji, oraz na tej podstawie oceniono wartość każdego rozwiązania. Ostatecznie podjęto próbę rozbudowania eksperymentu Monte Carlo

mającego dostarczyć informacji o jakości miar ewaluacji, w oparciu o ich podatność na sztucznie kontrolowane zmiany w danych. Wykorzystano go, aby porównać działania zaproponowanej miary i indeksu Jaccarda, przy różnych rodzajach zakłóceń.

2. Rozdział piąty opisuje przebieg i wyniki szeroko zakrojonego eksperymentu porównawczego jakości działania modułu percepcji MOD działającego na danych wizyjnych, sztucznie degradowanych na cztery różne sposoby o równomiernie stopniowanych poziomach natężenia. Jest to dość typowy przykład problemów i wyzwań jakie rodzi testowanie modułów percepcji bazujących na ogromnej ilości danych o dużej złożoności, od których oczekuje się uogólnienia szczegółowych wyników. Rezultatem eksperymentu było nie tylko zbadanie wpływu na wydajność analizowanego systemu wprowadzonych do danych zniekształceń, ale również potwierdzenie, że miary zaproponowane w rozdziale 4. sprawdzają się w warunkach złożonej analizy na masową skalę.

Kolejnym celem było zaproponowanie algorytmu tworzenia układu eksperymentalnego, uwzględniającego nie tylko rozróżnienie grup danych ze względu na warunki dostrzegalne gołym okiem takie jak pogoda, poziom urbanizacji i pora dnia, ale również ze względu naturę eksperymentu badającego wpływ zniekształceń wprowadzonych do danych wejściowych na wyniki ewaluacji. W tym celu zaprojektowano metodę klasteryzacji opartą o algorytm k-średnich i wykorzystującą rozkłady jakości sekwencji w scenariuszu jako reprezentacyjny obiekt, który można poddać klasteryzacji z wykorzystaniem metryki Wassersteina. Aby dopełnić algorytm, zaproponowano dwa podejścia do wyboru optymalnej liczby grup, uwzględniając późniejszą potrzebę przeprowadzenia testów ANOVA.

Ostatecznie, po uzyskaniu grup wyizolowanych ze względu na warunki naturalnie występujące w scenie oraz na kształt rozkładu oryginalnej jakości, dokonano analizy statystycznej wyników modułu percepcji pracujących na zdegradowanych danych wizyjnych. Skutkiem zastosowania zaproponowanego podziału na grupy z wykorzystaniem metryki Wassersteina była możliwość odizolowania części wariacji nie wynikającej z badanego w eksperymencie czynnika, co pozwoliło dokładniej wyznaczyć granice istotności różnic w parametrach badanych metodą statystyczną.

Tak uzyskane wyniki poddano dodatkowej obróbce topologicznej, ponownie postulując, że wartość jaką wniesie zaprzęgnięcie do pracy metryki Wassersteina leży w możliwości analizy całego kształtu rozkładu danych. Przeanalizowano jak zmie-

nia się odległość rozkładów jakości względem oryginalnych rozkładów we wszystkich badanych grupach. Umożliwiło to uzupełnienie wnioskowania statystycznego o dodatkowe spojrzenie nieparametryczne na kształty rozkładów uzyskanych po wprowadzeniu sztucznego zaburzenia, dostarczyło dodatkowych dowodów na nieodrzućenie hipotezy zerowej w brzegowych przypadkach i otworzyło możliwość nowego sposobu wizualizacji wyników eksperymentu.

3. W rozdziale szóstym skupiono się na analizie i charakteryzacji struktury samych danych testowych. Aby móc formalnie oceniać jakość danych, jakie wykorzystuje się do rozwoju oprogramowania (zarówno trening, testowanie jak i walidacja), przedstawiono sposób analizy kolejno podobieństwa samych trajektorii, scenariuszy jako zbioru trajektorii zebranych w jednej spójnej sytuacji drogowej oraz całych zbiorów scenariuszy.

To co zostało dokonane to przetestowanie metryk opisujących podobieństwo trajektorii sprowadzonych do działania na ograniczonej przestrzeni stanów, składającej się na obszar zasięgu sensorów wokół samochodu testowego. Wykorzystano do tego wybrane zestawy trajektorii z publicznego zbioru danych.

Kolejnym osiągnięciem było rozszerzenie koncepcji siatki zajętości do histogramu i rozkładu zajętości generowanego przez zbiory trajektorii, stanowiącego unikalną metodę reprezentacji złożonej sytuacji na drodze. Uwzględnia ona to, że na scenariusz składa się czynny udział samochodu testowego i wszystkich obiektów obecnych, poruszających się i okupujących przestrzeń wokół niego. Takie podejście pozwoliło zaproponować metodę wizualizacji zawartości scenariusza i sposobu na formalne porównywanie scenariuszy testowych za pomocą metryki Wassersteina. Dokonano eksperymentu porównawczego na wybranym zbiorze scenariuszy, który potwierdza zbieżność z intuicyjnym podejściem do rozróżniania scenariuszy.

Powyższy eksperyment pozwolił zaproponować metodologię porównawczą dla dużych zbiorów scenariuszy, wykorzystującą metrykę Wassersteina, algorytm klastrowania k-średnich i indeks Dunna. W rezultacie uzyskano możliwość wykorzystania reprezentacji scenariuszy przez rozkład zajętości do porównywania dwóch dużych zbiorów scenariuszy dzięki zaproponowanym wzorom współczynników różnorodności. Umożliwiają one opis podwójnego procesu klastrowania, który pozwala odpowiedzieć na pytanie, czy nowy zbiór danych wnosi istotną informację do już zebranych danych i jakiego rodzaju ona jest. Aby zaprezentować tą koncepcję podjęto się

porównania za jej pomocą dwóch zbiorów danych o rozmiarach 150 i 50 scenariuszy testowych.

Indeks Jaccarda ocenia dane związane z analizą rozpoznania jakości obiektów w sposób binarny na poziomie pojedynczego piksela, dlatego rozkłady jakości oparte o jego wyniki są mocno zbliżone do rozkładu normalnego. Analiza jakości, jak należałoby na podstawie struktury całej powyższej pracy stwierdzić, jest wielopoziomowa, od podstawowego porównania czterech współrzędnych opisujących prostokąt, poprzez ich sekwencje, dalej - zbiory sekwencji w scenariuszach pobranych w różnych warunkach, aż po badanie układów takich zbiorów. Każdy kolejny poziom analizy, korzystając z wyników na niższym poziomie, przeznaczony jest do uwzględnienia innego aspektu, który powinien mieć konkretny wpływ na wynik końcowy całego procesu. Koncentrując ocenę na każdym z tych poziomów jedynie na bezrefleksyjnym uśrednianiu, które mocno determinuje pozycje obserwacji lepszych i gorszych od przeciętnej w rozkładzie, co prawda otrzymamy pojęcie o tym, jaka jest oczekiwana jakość systemu, ale bardziej szczegółowe wnioskowanie na tej podstawie byłoby mocno wątpliwe. Tyczy się to również, analizy wysokopoziomowej, gdzie przechodzimy do korzystania z klasycznych rozwiązań testowania statystycznego. ANOVA analizuje parametry rozkładu, a przede wszystkim średnią. Mimo swojej uniwersalności i wielu innych zalet, sposób w jaki interpretuje dane sprowadzany jest poprzez centralne twierdzenie graniczne do rozkładu normalnego, czyli analizy parametrycznej. Zaproponowane rozwiązanie równoległe stara się tą analizę rozszerzyć na wszystkich poziomach. Zaczynając od podstaw, proponuje się miarę bardziej stanowczo wprowadzającą rozłam między jakością akceptowalną i nieakceptowalną. Następnie wyraźnie rozdziela się jakość sekwencji spełniających istotne kryteria bezpieczeństwa na drodze, oparte na potrzebie szybkiej reakcji od tych niewystarczająco sprawnie wskazujących nowe okoliczności. Ostatecznie analizując percepcję pojazdu, uwzględnia się jej zróżnicowaną skuteczność w różnych oczywistych i nieoczywistych warunkach. Te elementy mają wpływ na meta analizę jak widać w eksperymentach z rozdziału piątego. Kształt wykresu dla najważniejszej składowej - miary dystansu jest podstawą, na jakiej rozpinają się wyniki w eksperymentach, co widać w krzywiźnie uformowanej przez przedziały ufności stanowiące reprezentację wyników ANOVA. Rozumiejąc to, dostrzeżona została wartość w dopełnieniu tej analizy przez metrykę Wassersteina, która ma potencjał by w sposób nieparametryczny przestudiować tranzycję masy z wypuklonych wcześniej krytycznych poziomów akceptacji. Takie podejście, nie opierając swojej interpretacji wyłącznie na uśrednianiu wyniku ale na równoważnym traktowaniu całego kształtu rozkładów danych, pozwala na ich dokładniejsze porównanie, skutecznie wykorzystując fakt, że od podstaw opracowano algorytm ewalu-

acji z intencją, by wyeksponować w nich najważniejsze wnioski z pośrednich poziomów analiz. Innymi słowy metodologia kładzie mocny nacisk na zadbanie o to by precyzyjnie analizować dane na bardzo podstawowym poziomie, z uwzględnieniem ich szerokiego kontekstu, unikanie przedwczesnych generalizacji, gubiących istotne informacje oraz czerpanie z tego bezpośrednich korzyści na bardzo wysokim poziomie procesu ewaluacji, gdzie nieubłagane, mimo szeregu analiz zmierzających do generalizacji zbliża się moment, gdzie oczekiwane od badacza będzie syntetyzowanie konkretnych, dyskretnych wniosków aby podjąć istotne z perspektywy dużego przedsięwzięcia decyzje.

7.2. Dalsze badania

Jakkolwiek użyteczną jest możliwość precyzyjnej definicji co można, a czego nie można uznać za poprawne rozpoznanie, wiąże się to też z negatywnymi skutkami. W praktyce może stanowić wyzwanie ustalenie tego na linii klient - dostawca. Ponadto, różne kalibracje miar pomiędzy różnymi badaniami mogą utrudnić porównywanie wyników w dwóch różnych projektach, więc należy na to zwrócić szczególną uwagę. Naturalnie zmiany definicji podczas projektu mogą prowadzić do problemów z porównywaniem danych jakościowych w skali czasu, dlatego należy określić satysfakcjonującą definicję na początku, aby zadbać o jej niezmiennosc i uniknąć konieczności powtórnego przeliczania statystyk. Problemem, którego nie obejmuje wystarczająco analiza jakości zaprezentowana w pracy jest statystyka dotycząca rozpoznań fałszywie pozytywnych. Jest to bardzo istotna informacja, którą należałoby równolegle rozpatrywać z wynikami uzyskanymi na jej podstawie oraz określić jej wpływ na ostateczną ewaluację. Zagadnienie to choć poruszone w pracy nie znajduje w niej rozstrzygnięcia. Wynika to z potrzeby wyjścia poza analizę danych referencyjnych co rodzi potrzebę alternatywnego podejścia do automatyzacji tego procesu. Kontynuacja badań powinna zatem zawierać realizację pomysłu automatyzacji analizy rozpoznań fałszywie pozytywnych w oparciu, o klasyfikator ich poziomu szkodliwości. Taki klasyfikator mógłby wykorzystać metrykę Wasssersteina do ekstrapolowania treści obrazu wyznaczonego przez dane referencyjne. Następnie na podstawie porównania barycentrum wyznaczonego przez fałszywie pozytywny zbiór detekcji dotyczących danego obiektu określić czy zakorzenione rozpoznanie fałszywie pozytywne oznaczone przez algorytm opisany w sekcji 4.3.1 jest istotnie szkodliwe czy może wynika z nieprecyzyjnego lub niespełniającego praktycznych wymagań etykietowania. Wymagałoby to wykorzystania surowych danych wizyjnych bezpośrednio w procesie ewaluacyjnym co znacznie podnosi skalę wyzwania dla algorytmu przetwarzającego duże ilości takich danych. Tak zauto-

matyzowana analiza mogłaby znacznie poprawić wyznacznik precyzji modułu percepcji. Korekta tej wartości byłaby uzyskana poprzez obniżenie wartości mianownika o ilość rozpoznań fałszywie pozytywnych sklasyfikowanych jako nieszkodliwe.

Co się tyczy natomiast charakteryzacji danych zebranych podczas logowania scenariuszy testowych praca skupia się na deskrypcji własności pozycji i dynamiki obiektów w polu widzenia. Należałoby rozszerzyć taką charakterystykę o własności niemożliwe do odczytania przez samą analizę trajektorii. Mowa tu na przykład o prędkości obiektów, o kierunkowskazach (negocjacji zmiany pasa), treści znaków drogowych oraz interakcji między obiektami obecnymi w tym samym momencie w scenariuszu. Analiza zajętości siatki nie dostarcza części tych informacji lub może je niewystarczająco reprezentować. Częściowym rozwiązaniem tego mogłaby być trójwymiarowa mapa termiczna rozpinająca dotychczasowe warstwy mówiące o zajętości otoczenia na trzeciej osi utożsamionej z czasem. Wartości pól powinny wyznaczać natężenie parametrów poszczególnych obiektów (ciągłe - prędkość, a dyskretne - kierunkowskaz lub typ znaku). Choć możliwość manualnej interpretacji takich macierzy przez człowieka byłaby ograniczona to wciąż stanowiłyby obiekt, który można poddać analizie odległości Wassersteina, a zatem i klastrowaniu dającym informację o różnorodności zbiorów scenariuszy.

Bibliografia

- [1] M. Komorkiewicz, A. Chin, P. Skruch, and M. Szelest, “Big data challenges in current and next generation automated vehicle development,” *Recenzja w toku*.
- [2] J. Vargas, S. Alswiss, O. Toker, R. Razdan, and J. Santos, “An overview of autonomous vehicles sensors and their vulnerability to weather conditions,” *Sensors*, vol. 21, no. 16, p. 5397, 2021.
- [3] APTIV - MOBILITY INSIDER, “How machine learning improves perception.” <https://www.aptiv.com/en/insights/article/how-machine-learning-improves-perception>, 2020.
- [4] APTIV - MOBILITY INSIDER, “What is sensor fusion?” <https://www.aptiv.com/en/insights/article/what-is-sensor-fusion>, 2020.
- [5] APTIV - MOBILITY INSIDER, “What is adas?” <https://www.aptiv.com/en/insights/article/what-is-adas>, 2020.
- [6] APTIV - MOBILITY INSIDER, “What is active safety?” <https://www.aptiv.com/en/insights/article/what-is-active-safety>, 2020.
- [7] F. Duchoň, P. Hubinský, J. Hanzel, A. Babinec, and M. Tölgyessy, “Intelligent vehicles as the robotic applications,” *Procedia Engineering*, vol. 48, pp. 105–114, 2012. Modelling of Mechanical and Mechatronics Systems.
- [8] S. Shalev-Shwartz, S. Shammah, and A. Shashua, “On a formal model of safe and scalable self-driving cars,” *CoRR*, vol. abs/1708.06374, 2017.
- [9] N. Pankiewicz, W. Turlej, M. Orłowski, and T. Wrona, “Highway pilot training from demonstration,” in *2021 25th International Conference on Methods and Models in Automation and Robotics (MMAR)*, pp. 109–114, 2021.

- [10] S. Shalev-Shwartz, S. Shammah, and A. Shashua, “Safe, multi-agent, reinforcement learning for autonomous driving,” *CoRR*, vol. abs/1610.03295, 2016.
- [11] M. Uříčář, D. Hurych, P. Krizek, and S. Yogamani, “Challenges in designing datasets and validation for autonomous driving,” 01 2019.
- [12] “Road vehicles — Safety of the intended functionality,” standard, International Organization for Standardization, 2019.
- [13] M. Rabe, S. Milz, and P. Mader, “Development methodologies for safety critical machine learning applications in the automotive domain: A survey,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 129–141, 2021.
- [14] P. Kowalczyk, J. Izydorczyk, and M. Szelest, “Evaluation methodology for object detection and tracking in bounding box based perception modules,” *Electronics*, vol. 11, no. 8, 2022.
- [15] P. Kowalczyk, P. Bugiel, J. Izydorczyk, and M. Szelest, *Klastrowanie scenariuszy testowych za pomocą metryki Wassersteina i analiza jakości rozpoznań obiektów w optycznie zdegradowanych filmach*. Gliwice, Polska: Elektronika, Telekomunikacja, Mobilność, Wydawnictwo Politechniki Śląskiej, 2021.
- [16] P. Kowalczyk, P. Bugiel, M. Szelest, and J. Izydorczyk, “Fault injection in optical path - detection quality degradation analysis with wasserstein distance,” in *2021 25th International Conference on Methods and Models in Automation and Robotics (MMAR)*, pp. 121–126, 2021.
- [17] P. Skruch, M. Szelest, and P. Kowalczyk, “An approach for evaluating the completeness of the test scenarios for the vehicle environmental perception-based systems,” in *2021 25th International Conference on Methods and Models in Automation and Robotics (MMAR)*, pp. 133–138, 2021.
- [18] P. Kowalczyk, M. Komorkiewicz, P. Skruch, and M. Szelest, “Efficient characterization method for big automotive datasets used for perception system development and verification,” *IEEE Access*, pp. 1–1, 2022.
- [19] J. S. Brodsky, “Autonomous vehicle regulation: How an uncertain legal landscape may hit the brakes on self-driving cars,” *Berkeley Technology Law Journal*, vol. 31, p. 851, 2016.

- [20] L. H. Collingwood, “Privacy implications and liability issues of autonomous vehicles,” *Information & Communications Technology Law*, vol. 26, pp. 32 – 45, 2017.
- [21] D. Danks and A. J. London, “Regulating autonomous systems: Beyond standards,” *IEEE Intelligent Systems*, vol. 32, pp. 88–91, 2017.
- [22] “2016 Fatal Motor Vehicle Crashes: Overview, National Highway Transportation Safety Administration (NHTSA).” <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812456>.
- [23] J. Wei, J. M. Dolan, and B. Litkouhi, “Autonomous vehicle social behavior for highway entrance ramp management,” in *2013 IEEE Intelligent Vehicles Symposium (IV)*, pp. 201–207, 2013.
- [24] N. Lichtlé, E. Vinitsky, G. Gunter, A. Velu, and A. M. Bayen, “Fuel consumption reduction of multi-lane road networks using decentralized mixed-autonomy control,” in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 2068–2073, 2021.
- [25] Dz.U. 1997 nr 98 poz. 602, “Ustawa z dnia 20 czerwca 1997 r. - prawo o ruchu drogowym,” 1997. z późniejszymi zmianami.
- [26] “Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2019/2144 z dnia 27 listopada 2019 r. w sprawie wymogów dotyczących homologacji typu pojazdów silnikowych i ich przyczep oraz układów, komponentów i oddzielnych zespołów technicznych przeznaczonych do tych pojazdów, w odniesieniu do ich ogólnego bezpieczeństwa oraz ochrony osób znajdujących się w pojeździe i niechronionych uczestników ruchu drogowego, zmieniające rozporządzenie Parlamentu Europejskiego i Rady (UE) 2018/858 oraz uchylające rozporządzenia Parlamentu Europejskiego i Rady (WE) nr 78/2009, (WE) nr 79/2009 i (WE) nr 661/2009 oraz rozporządzenia Komisji (WE) nr 631/2009, (UE) nr 406/2010, (UE) nr 672/2010, (UE) nr 1003/2010, (UE) nr 1005/2010, (UE) nr 1008/2010, (UE) nr 1009/2010, (UE) nr 19/2011, (UE) nr 109/2011, (UE) nr 458/2011, (UE) nr 65/2012, (UE) nr 130/2012, (UE) nr 347/2012, (UE) nr 351/2012, (UE) nr 1230/2012 i (UE) 2015/166,” listopad 2019.
- [27] H. L. Bloecher, J. Dickmann, and M. Andres, “Automotive active safety & comfort functions using radar,” in *2009 IEEE International Conference on Ultra-Wideband*, pp. 490–494, 2009.

- [28] “Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles,” *SAE International J3016*, 2021 (revision).
- [29] APTIV - MOBILITY INSIDER, “What is a driver-monitoring system?.” <https://www.aptiv.com/en/insights/article/what-is-a-driver-monitoring-system>, 2021.
- [30] APTIV - MOBILITY INSIDER, “On adas, autonomous driving, and vehicle safety technologies.” <https://www.rohm.com/blog/-/blog/id/8030373>, 2020.
- [31] “Tesla Model S.” https://www.tesla.com/pl_pl/models.
- [32] “Audi A6 Limousine.” <https://www.audi.pl/pl/web/pl/modele/a6/s6-limousine/wyposazenie.html>.
- [33] “Audi A8L.” <https://www.audi.pl/pl/web/pl/modele/a8/a8-l.html>.
- [34] “Frontowa Kamera.” <http://www.eurocarnews.com/media/pictorials/933/4278.jpg>.
- [35] V. T. B. Tram and M. Yoo, “Vehicle-to-vehicle distance estimation using a low-resolution camera based on visible light communications,” *IEEE Access*, vol. 6, pp. 4521–4527, 2018.
- [36] M. A. Haseeb, J. Guan, D. Ristic-Durrant, and A. Gräser, “Disnet: a novel method for distance estimation from monocular camera,” *10th Planning, Perception and Navigation for Intelligent Vehicles (PPNIV18), IROS*, 2018.
- [37] M. Gadringer, F. Maier, H. Schreiber, V. Makkapati, A. Gruber, M. Vorderderfler, D. Amschl, S. Metzner, H. Pflügl, W. Bösch, M. Horn, and M. Paulweber, “Radar target stimulation for automotive applications,” *IET Radar, Sonar & Navigation*, vol. 12, pp. 1096–1103, Sept. 2018.
- [38] “Short Range Radar – SRR320.” <https://www.continental-automotive.com/en-gl/2-Wheeler/Safe-Mobility/Sensors/Short-Range-Radar-SRR320>.
- [39] “Advanced Radar Sensor ARS540.” <https://www.continental-automotive.com/en-gl/Passenger-Cars/Autonomous-Mobility/Enablers/Radars/Long-Range-Radar/ARS540>.
- [40] P. Khomchuk, I. Stainvas, and I. Bilik, “Pedestrian motion direction estimation using simulated automotive MIMO radar,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, p. 1132–1145, Jun 2016.

- [41] F. Jiménez, J. E. Naranjo, and F. García, “An improved method to calculate the time-to-collision of two vehicles,” *International Journal of Intelligent Transportation Systems Research*, 2013.
- [42] “Tor testowy Aptiv.” www.auto-swiat.pl/wiadomosci/aktualnosci/najwiekszy-osrodek-testowy-na-swiecie-czy-wiesz-ze-jest-w-polsce/91t6kpxslajd-12.
- [43] S. Royo and M. Ballesta-Garcia, “An overview of lidar imaging systems for autonomous vehicles,” *Applied Sciences*, vol. 9, no. 19, 2019.
- [44] “Bosh Ultrasonic system generation 6.5.” <https://www.bosch-mobility-solutions.com/en/solutions/assistance-systems/ultrasonic-sensor-systems-ohw/>.
- [45] J. Tiete, F. Domínguez, B. d. Silva, L. Segers, K. Steenhaut, and A. Touhafi, “Soundcompass: a distributed mems microphone array-based sensor for sound source localization,” *Sensors*, vol. 14, no. 2, pp. 1918–1949, 2014.
- [46] K. Kodera, A. Itai, and H. Yasukawa, “Sound localization of approaching vehicles using uniform microphone array,” in *2007 IEEE Intelligent Transportation Systems Conference*, pp. 1054–1058, 2007.
- [47] Y. Jang, J. Kim, and J. Kim, “The development of the vehicle sound source localization system,” in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1241–1244, 2015.
- [48] P. Markiewicz, M. Długosz, and P. Skruch, “Review of tracking and object detection systems for advanced driver assistance and autonomous driving applications with focus on vulnerable road users sensing,” in *Trends in Advanced Intelligent Control, Optimization and Automation* (W. Mitkowski, J. Kacprzyk, K. Oprzędkiewicz, and P. Skruch, eds.), (Cham), pp. 224–237, Springer International Publishing, 2017.
- [49] K.-H. Chen, K.-S. Cheng, and J.-Y. He, “Perception module design of self-driving vehicles for navigation in people-rich environments,” in *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, pp. 1–2, 2020.
- [50] E. Romera, L. Bergasa, and R. Arroyo, “Can we unify monocular detectors for autonomous driving by using the pixel-wise semantic segmentation of CNNs?,” 06 2016.

- [51] Z.-H. Zhou, “A brief introduction to weakly supervised learning,” *National Science Review*, vol. 5, pp. 44–53, 08 2017.
- [52] L. Yang, G. Hu, Y. Song, G. Li, and L. Xie, “Intelligent video analysis: A pedestrian trajectory extraction method for the whole indoor space without blind areas,” *Computer Vision and Image Understanding*, vol. 196, p. 102968, 2020.
- [53] P. Koopman and M. Wagner, “Toward a framework for highly automated vehicle safety validation,” in *WCX World Congress Experience*, SAE International, apr 2018.
- [54] W. Zimmer, A. Rangesh, and M. Trivedi, “3D BAT: A Semi-Automatic, Web-based 3D Annotation Toolbox for Full-Surround, Multi-Modal Data Streams,” in *2019 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1816–1821, 2019.
- [55] “Automation systems and integration — Key performance indicators (KPIs) for manufacturing operations management — Part 1: Overview, concepts and terminology,” standard, International Organization for Standardization, 2014.
- [56] “Automation systems and integration — Key performance indicators (KPIs) for manufacturing operations management — Part 2: Definitions and descriptions, institution = International Organization for Standardization,” standard, 2014.
- [57] B. E. Baptiste Hemery, Hélène Laurent and C. Rosenberger, “Comparative study of localization metrics for the evaluation of image interpretation systems.,” *Journal of Electronic Imaging, SPIE and IS&T*, 2010.
- [58] B. Hemery, H. Laurent, and C. Rosenberger, “Comparative study of metrics for evaluation of object localisation by bounding boxes.,” in *Fourth International Conference on Image and Graphics (ICIG 2007)*, pp. 459–464, 2007.
- [59] D. Zhou, J. Fang, X. Song, C. Guan, J. Yin, Y. Dai, and R. Yang, “IoU Loss for 2D/3D Object Detection,” in *2019 International Conference on 3D Vision (3DV)*, pp. 85–94, 2019.
- [60] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, “A benchmark dataset and evaluation methodology for video object segmentation,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 724–732, 2016.

- [61] A. A. Taha, A. Hanbury, and O. A. J. del Toro, “A formal method for selecting evaluation metrics for image segmentation,” in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 932–936, 2014.
- [62] Z. Qin, Q. Li, H. Li, X. Dong, and Z. Ren, “Advanced intersection over union loss for visual tracking,” in *2019 Chinese Automation Congress (CAC)*, pp. 5869–5873, 2019.
- [63] N. Lazarevic-McManus, J. Renno, D. Makris, and G. Jones, “An object-based comparative methodology for motion detection based on the F-Measure,” *Computer Vision and Image Understanding*, vol. 111, no. 1, pp. 74–85, 2008. Special Issue on Intelligent Visual Surveillance (IEEE).
- [64] L. Cehovin, M. Kristan, and A. Leonardis, “Is my new tracker really better than yours?,” in *IEEE Winter Conference on Applications of Computer Vision*, pp. 540–547, 2014.
- [65] L. Cehovin, A. Leonardis, and M. Kristan, “Visual object tracking performance measures revisited,” *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1261–1274, 2016.
- [66] T. Nawaz and A. Cavallaro, “A protocol for evaluating video trackers under real-world conditions,” *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1354–1361, 2013.
- [67] M. Szczodrak, P. Dalka, and A. Czyżewski, “Performance evaluation of video object tracking algorithm in autonomous surveillance system,” in *2010 2nd International Conference on Information Technology, (2010 ICIT)*, pp. 31–34, 2010.
- [68] M. Berman, A. R. Triki, and M. B. Blaschko, “The lovasz-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4413–4421, 2018.
- [69] Lurong Shen, Xinsheng Huang, Yuzhuang Yan, and Shengjian Bai, “An improved mean-shift tracking algorithm with spatial-color feature and new similarity measure,” in *2011 International Conference on Multimedia Technology*, pp. 184–188, 2011.

- [70] Meng Zhang, Ken Chen, and Yun Yan Guo, “Online parameter based kalman filter precision evaluation method for video target tracking,” in *2011 International Conference on Multimedia Technology*, pp. 598–601, 2011.
- [71] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang, “Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 319–336, 2009.
- [72] J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbelaez, A. Sorkine-Hornung, and L. V. Gool, “The 2017 DAVIS challenge on video object segmentation,” *CoRR*, vol. abs/1704.00675, 2017.
- [73] Y. Fang and X. Fan, “Performance evaluation for ir small target tracking algorithm,” in *2011 Sixth International Conference on Image and Graphics*, pp. 749–753, 2011.
- [74] J. Tu, A. Del Amo, Y. Xu, L. Guari, M. Chang, and T. Sebastian, “A fuzzy bounding box merging technique for moving object detection,” in *2012 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)*, pp. 1–6, 2012.
- [75] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, “Visual tracking: An experimental survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442–1468, 2014.
- [76] Chee-Yee Chong, “Problem characterization in tracking/fusion algorithm evaluation,” *IEEE Aerospace and Electronic Systems Magazine*, vol. 16, no. 7, pp. 12–17, 2001.
- [77] G. Liu, S. Liu, M. Lu, and Z. Pan, “Effects of improper ground truth on target tracking performance evaluation in benchmark,” in *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pp. 261–266, 2017.
- [78] J. He, S. M. Erfani, X. Ma, J. Bailey, Y. Chi, and X. Hua, “Alpha-IoU: A Family of Power Intersection over Union Losses for Bounding Box Regression,” *CoRR*, vol. abs/2110.13675, 2021.

- [79] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression,” *CoRR*, vol. abs/1911.08287, 2019.
- [80] V. V. Aroulanandam, T. P. Latchoumi, B. Bhavya, and S. S. Sultana, “Object detection in convolution neural networks using iterative refinements,” *architecture*, vol. 15, p. 17, 2019.
- [81] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 658–666, 2019.
- [82] M. Everingham, L. V. Gool, C. K. I. Williams, J. M. Winn, and A. Zisserman, “The Pascal Visual Object Classes (VOC) Challenge,” *International Journal of Computer Vision*, vol. 88, pp. 303–338, 2009.
- [83] D. Riahi and G.-A. Bilodeau, “Online multi-object tracking by detection based on generative appearance models,” *Computer Vision and Image Understanding*, vol. 152, pp. 88–102, 2016.
- [84] B. Ristic, B. Vo, and D. Clark, “Performance evaluation of multi-target tracking using the OSPA metric,” in *2010 13th International Conference on Information Fusion*, pp. 1–7, 2010.
- [85] B. Ristic, B. Vo, D. Clark, and B. Vo, “A metric for performance evaluation of multi-target tracking algorithms,” *IEEE Transactions on Signal Processing*, vol. 59, no. 7, pp. 3452–3457, 2011.
- [86] X. Shi, F. Yang, F. Tong, and H. Lian, “A comprehensive performance metric for evaluation of multi-target tracking algorithms,” in *2017 3rd International Conference on Information Management (ICIM)*, pp. 373–377, 2017.
- [87] M. Kristan, J. Matas, A. Leonardis, T. Vojíř, R. Pflugfelder, G. Fernández, G. Nebhay, F. Porikli, and L. Čehovin, “A novel performance evaluation methodology for single-target trackers,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 11, pp. 2137–2155, 2016.

- [88] A. Tokta and A. K. Hocaoglu, “A track to track association algorithm based on weighted state correlation similarity,” in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, pp. 1–4, 2018.
- [89] P. Pan, F. Porikli, and D. Schonfeld, “A new method for tracking performance evaluation based on a reflective model and perturbation analysis,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3529–3532, 2009.
- [90] E. Loutas, N. Nikolaidis, and I. Pitas, “Evaluation of tracking reliability metrics based on information theory and normalized correlation,” in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 4, pp. 653–656 Vol.4, 2004.
- [91] I. Batyrshin, “Constructing time series shape association measures: Minkowski distance and data standardization,” in *2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence*, pp. 204–212, 2013.
- [92] S. Mori, K. Chang, and C. Chong, “Comparison of track fusion rules and track association metrics,” in *2012 15th International Conference on Information Fusion*, pp. 1996–2003, 2012.
- [93] D. Roth, E. Koller-Meier, D. Rowe, T. B. Moeslund, and L. Van Gool, “Event-based tracking evaluation metric,” in *2008 IEEE Workshop on Motion and Video Computing*, pp. 1–8, 2008.
- [94] T. Vu and R. Evans, “A new performance metric for multiple target tracking based on optimal subpattern assignment,” in *17th International Conference on Information Fusion (FUSION)*, pp. 1–8, 2014.
- [95] Y. Okafuji, C. D. Mole, N. Merat, T. Fukao, Y. Yokokohji, H. Inou, and R. M. Wilkie, “Steering bends and changing lanes: The impact of optic flow and road edges on two point steering control,” *Journal of Vision*, vol. 18, pp. 14–14, 09 2018.
- [96] G. Guo and W. Yue, “Sampled-data cooperative adaptive cruise control of vehicles with sensor failures,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2404–2418, 2014.

- [97] S. Hasirlioglu and A. Riener, “Challenges in object detection under rainy weather conditions,” in *Intelligent Transport Systems, From Research and Development to the Market Uptake* (J. C. Ferreira, A. L. Martins, and V. Monteiro, eds.), (Cham), pp. 53–65, Springer International Publishing, 2019.
- [98] L. Tang, Y. Shi, Q. He, A. W. Sadek, and C. Qiao, “Performance test of autonomous vehicle lidar sensors under different weather conditions,” *Transportation Research Record*, vol. 2674, no. 1, pp. 319–329, 2020.
- [99] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Gläser, F. Timm, W. Wiesbeck, and K. Dietmayer, “Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2021.
- [100] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “CARLA: An open urban driving simulator,” in *Proceedings of the 1st Annual Conference on Robot Learning*, pp. 1–16, 2017.
- [101] T. Sulkowski, P. Bugiel, and J. Izydorczyk, “In search of the ultimate autonomous driving simulator,” in *2018 International Conference on Signals and Electronic Systems (ICSES)*, pp. 252–256, 2018.
- [102] B. Kim, H. Kim, K. Kim, S. Kim, and J. Kim, “Learning not to learn: Training deep neural networks with biased data,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [103] R. Vilalta and Y. Drissi, “A perspective view and survey of meta-learning,” *Artificial Intelligence Review*, vol. 18, no. 2, pp. 77–95, 2002.
- [104] P. Skruch, R. Dlugosz, K. Kogut, P. Markiewicz, D. Sasin, and M. Rozewicz, “The simulation strategy and its realization in the development process of active safety and advanced driver assistance systems,” *SAE Technical Paper 2015-01-140*, 2015.
- [105] A. Kalousis, J. Gama, and M. Hilario, “On data and algorithms: Understanding inductive performance,” *Machine Learning*, vol. 54, no. 3, pp. 275–312, 2004.
- [106] N. Macia, E. Bernado-Mansilla, A. Orriols-Puig, and T. Ho, “Learner excellence biased by data set selection: A case for data characterisation and artificial data sets,” *Pattern Recognition*, vol. 46, no. 3, pp. 1054–1066, 2013.

- [107] R. Akase and Y. Okada, “Web based time-tunnel: An interactive multidimensional data visualization tool using genetic algorithm,” in *2015 19th International Conference on Information Visualisation*, pp. 95–100, 2015.
- [108] P. Volland, “Spatio-temporal visualization of automotive sensor data — a conceptual and implementational framework,” in *2017 IEEE Region 10 Symposium (TEN-SYMP)*, pp. 1–5, 2017.
- [109] J. Heinrich and D. Weiskopf, “Parallel coordinates for multidimensional data visualization: Basic concepts,” *Computing in Science Engineering*, vol. 17, no. 3, pp. 70–76, 2015.
- [110] R. Smeulders and A. Heijs, “Interactive visualization of high dimensional marketing data in the financial industry,” in *Ninth International Conference on Information Visualisation (IV’05)*, pp. 814–817, 2005.
- [111] R. Khlebnikov, B. Kainz, M. Steinberger, and D. Schmalstieg, “Noise-based volume rendering for the visualization of multivariate volumetric data,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2926–2935, 2013.
- [112] Z. Aung, K. Sim, and W. S. Ng, “Traj Align: A Method for Precise Matching of 3-D Trajectories,” in *2010 20th International Conference on Pattern Recognition*, pp. 3818–3821, 2010.
- [113] S. B. Needleman and C. D. Wunsch, “A general method applicable to the search for similarities in the amino acid sequence of two proteins,” *J. Mol. Biol.*, vol. 48, pp. 443–453, 1970.
- [114] H. Ma and M. Wang, “Trajectory similarity metric based on vector fields on manifold and multi-scale space partition,” in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 1600–1604, 2016.
- [115] X. Li, Z. Liu, Z. Liu, R. W. Liu, and Z. Feng, “Spatio-temporal vessel trajectory smoothing based on trajectory similarity and two-dimensional wavelet transform,” in *2019 5th International Conference on Transportation Information and Safety (ICTIS)*, pp. 1500–1505, 2019.

- [116] C. Jiang, Y. Zhou, and C. Zhang, "Similarity measurement of spatiotemporal trajectories considering semantic features," in *2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA)*, pp. 423–426, 2021.
- [117] B. Chen, Y. Liu, and W. Shi, "Research on trajectory similarity matching model based on spatiotemporal trajectory big data," in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, vol. 1, pp. 2208–2215, 2020.
- [118] M. Riyadh, N. Mustapha, and D. Riyadh, "Review of trajectories similarity measures in mining algorithms," in *2018 Al-Mansour International Conference on New Trends in Computing, Communication, and Information Technology (NTCCIT)*, pp. 36–40, 2018.
- [119] A. Croitoru, P. Agouris, and A. Stefanidis, "3D trajectory matching by pose normalization," *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, pp. 153–162, 01 2005.
- [120] J. Urain and J. Peters, "Generalized multiple correlation coefficient as a similarity measurement between trajectories," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1363–1369, 2019.
- [121] H. Fashandi and A. M. E. Moghaddam, "A new rotation invariant similarity measure for trajectories," in *2005 International Symposium on Computational Intelligence in Robotics and Automation*, pp. 631–634, 2005.
- [122] M. Vlachos, G. Kollios, and D. Gunopulos, "Discovering similar multidimensional trajectories," in *Proceedings 18th International Conference on Data Engineering*, pp. 673–684, 2002.
- [123] G. Monge, *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale, 1781.
- [124] S. Peleg, M. Werman, and H. Rom, "A unified approach to the change of resolution: space and gray-level," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 739–742, 1989.
- [125] Y. Rubner, C. Tomasi, and L. Guibas, "A metric for distributions with applications to image databases," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pp. 59–66, 1998.

- [126] L. N. Vaserstein, “Markov processes over denumerable products of spaces describing large system of automata,” vol. 5, p. 64–72, 1969.
- [127] L. V. Kantorovich, “The mathematical method of production planning and organization,” vol. 6, pp. 363–422, 1939.
- [128] J. Xu, L. Luo, C. Deng, and H. Huang, “Multi-Level Metric Learning via Smoothed Wasserstein Distance,” in *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI’18*, p. 2919–2925, AAAI Press, 2018.
- [129] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 1 ed., 2000.
- [130] M. Cuturi and A. Doucet, “Fast computation of wasserstein barycenters.” <https://arxiv.org/abs/1310.4375>, 2013.
- [131] J. Ebert, V. Spokoiny, and A. Suvorikova, “Construction of Non-asymptotic Confidence Sets in 2-Wasserstein Space,” 2017.
- [132] Y. Zeng, X. Fu, L. Gao, J. Zhu, H. Li, and Y. Li, “Robust multivehicle tracking with wasserstein association metric in surveillance videos,” *IEEE Access*, vol. 8, pp. 47863–47876, 2020.
- [133] J. Kline, “Properties of the d-dimensional earth mover’s problem,” *Discrete Applied Mathematics*, vol. 265, pp. 128–141, 2019.
- [134] Z. Hu, H. Xue, Q. Zhang, J. Gao, N. Zhang, S. Zou, Y. Teng, X. Liu, Y. Yang, D. Liang, X. Zhu, and H. Zheng, “DPIR-Net: Direct PET Image Reconstruction Based on the Wasserstein Generative Adversarial Network,” *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 35–43, 2021.
- [135] K. Lei, M. Mardani, J. M. Pauly, and S. S. Vasanawala, “Wasserstein GANs for MR Imaging: From Paired to Unpaired Training,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 105–115, 2021.
- [136] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, “Visual tracking: An experimental survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1442–1468, 2014.

- [137] R. S. Jurecki and T. L. Stańczyk, “Analyzing driver response times for pedestrian intrusions in crash-imminent situations,” in *2018 XI International Science-Technical Conference Automotive Safety*, pp. 1–7, 2018.
- [138] “IEEE Standard for Motor Vehicle Event Data Recorder (MVEDR),” *IEEE Std 1616-2004*, pp. 1–171, 2005.
- [139] “Relacje wartości długości drogi hamowania i drogi zatrzymania dla różnych warunków ruchu samochodu,” *Autobusy: technika, eksploatacja, systemy transportowe*, no. 6, 2010.
- [140] D. Huttenlocher and W. Rucklidge, “A multi-resolution technique for comparing images using the Hausdorff distance,” *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 705–706, 1993.
- [141] E. d’Angelo, S. Herbin, and M. Ratiéville, “Robin challenge evaluation principles and metrics,” 2006.
- [142] W. K. Pratt, O. D. Faugeras, and A. Gagalowicz, “Visual discrimination of stochastic texture fields,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 11, pp. 796–804, 1978.
- [143] G. Bonaccorso, “Mastering machine learning algorithms.” Packt Publishing, 2018.
- [144] W. Cunningham, “A network simplex method,” *Math Program*, vol. 11, pp. 105–116, 12 1976.
- [145] N. Bonneel, M. van de Panne, S. Paris, and W. Heidrich, “Displacement interpolation using lagrangian mass transport,” *ACM Trans. Graph.*, vol. 30, p. 1–12, Dec. 2011.
- [146] D. Arthur and S. Vassilvitskii, “K-means++: The advantages of careful seeding,” vol. 8, pp. 1027–1035, 01 2007.
- [147] H. Mahi, N. Farhi, K. Labed, and D. Benhamed, “The silhouette index and the k-harmonic means algorithm for multispectral satellite images clustering,” in *2018 International Conference on Applied Smart Systems (ICASS)*, pp. 1–6, 2018.
- [148] H. Mahi, N. Farhi, K. Labed, and D. Benhamed, “The silhouette index and the k-harmonic means algorithm for multispectral satellite images clustering,” in *2018 International Conference on Applied Smart Systems (ICASS)*, pp. 1–6, 2018.

- [149] M. Nixon and A. S. Aguado, *Feature Extraction & Image Processing for Computer Vision, Third Edition*. USA: Academic Press, Inc., 3rd ed., 2012.
- [150] H. Liu, *Comparing Welch's ANOVA, a Kruskal-Wallis Test, and Traditional ANOVA in Case of Heterogeneity of Variance*. Virginia Commonwealth University, 2015.
- [151] Minitab, LLC, "Minitab 18." <https://www.minitab.com>.
- [152] D. C. Sauder and C. E. DeMars, "An updated recommendation for multiple comparisons," *Advances in Methods and Practices in Psychological Science*, vol. 2, no. 1, pp. 26–44, 2019.
- [153] J. L. Gastwirth, Y. R. Gel, and W. Miao, "The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice," *Statistical Science*, vol. 24, Aug 2009.
- [154] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, 01 2008.
- [155] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset," *International Journal of Robotics Research (IJRR)*, 2013.
- [156] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nusenes: A multimodal dataset for autonomous driving," *arXiv preprint arXiv:1903.11027*, 2019.
- [157] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn, T. Fernandez, M. Jänicke, S. Mirashi, C. Savani, M. Sturm, O. Vorobiov, M. Oelker, S. Garreis, and P. Schubert, "A2D2: audi autonomous driving dataset," *CoRR*, vol. abs/2004.06320, 2020.
- [158] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. Qi, Y. Zhou, Z. Yang, A. Chouard, P. Sun, J. Ngiam, V. Vasudevan, A. McCauley, J. Shlens, and D. Anguelov, "Large Scale Interactive Motion Forecasting for Autonomous Driving : The Waymo Open Motion Dataset," *CoRR*, vol. abs/2104.10133, 2021.

-
- [159] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, “BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning,” 2020.
- [160] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, and R. Yang, “The apolloscape open dataset for autonomous driving and its application,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2702–2719, 2020.
- [161] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” 2016.
- [162] J. C. Dunn†, “Well-separated clusters and optimal fuzzy partitions,” *Journal of Cybernetics*, vol. 4, no. 1, pp. 95–104, 1974.