



Politechnika Wrocławska

Wydział Informatyki i Zarządzania

Kierunek studiów: Informatyka

Specjalność: Danologia

Praca dyplomowa – magisterska

**ANALIZA OTOCZENIA POJAZDU NA PRZYKŁADZIE
LOKALIZACJI PALET W ŚRODOWISKU WÓZKA
WIDŁOWEGO**

Maciej Gruszczyński

słowa kluczowe:

detekcja, obraz, palety, wózek widłowy

krótkie streszczenie:

W ramach pracy wybrano, opracowano, zespolono i zbadano metody detekcji obiektu na obrazie na przykładzie palet w środowisku wózka widłowego.

Opiekun pracy diplomowej	dr inż. Jerzy Sas
	Tytuł/stopień naukowy/imię i nazwisko	ocena	podpis
Ostateczna ocena za pracę dyplomową			
Przewodniczący Komisji egzaminu diplomowego	
	Tytuł/stopień naukowy/imię i nazwisko	ocena	podpis

*Do celów archiwalnych pracę dyplomową zakwalifikowano do:**

- a) kategorii A (akta wieczyste)
- b) kategorii BE 50 (po 50 latach podlegające ekspertyzie)

* niepotrzebne skreślić

pieczętka wydziałowa

Wrocław, rok 2020

SPISTREŚCI

Spis rysunków	3
Spis tabel	5
Streszczenie	6
Wstęp	8
Opis problemu	8
Cel pracy	8
Struktura pracy	9
1. Opis problemu	10
1.1. Umiejscowienie w szerszej rzeczywistości	10
1.2. Szczegóły problemu detekcji palet	11
1.2.1. Cechy wykrywanego obiektu	11
1.2.2. Wpływ warunków zewnętrznych	12
1.2.3. Pozycja kamery	13
2. Przegląd literaturowy	16
2.1. Rozwiązania dotyczące bezpośrednio detekcji palet	16
2.1.1. Wprowadzenie	16
2.1.2. Metody oparte na binaryzacji	16
2.1.3. Metody oparte na ekstrakcji cech z obrazów w skali szarości	19
2.1.4. Ewaluacja i wyniki	23
2.1.5. Podsumowanie	25
2.2. Odszumianie obrazów	27
2.2.1. Wprowadzenie	27
2.2.2. Przegląd istniejących metod odszumiania	27
2.2.3. Podsumowanie w odniesieniu do problemu detekcji palet	31
2.3. Wygładzanie detekcji w sekwencji wideo	31
2.3.1. Wprowadzenie	31
2.3.2. Prosta estymacja	32
2.3.3. Filtry alfa	32
2.3.4. Filtry Kalman'a	33
2.3.5. Filtr Kalman'a w śledzeniu obiektów	35
2.3.6. Podsumowanie	36
2.4. Podsumowanie przeglądu	36
3. Opis badanej metody	38
3.1. Architektura wysokiego poziomu	38
3.2. Technika okna przesuwnego	38
3.3. Detekcja w oparciu o barwę	40
3.4. Detekcja w oparciu o gradient	42
3.5. Zespół metod detekcji	44
3.6. Ewaluacja	46
4. Badania i eksperymenty	49

4.1.	Zbiór danych	49
4.1.1.	Pozyskanie i cechy	49
4.1.2.	Etykietowanie	49
4.2.	Badania metod składowych	49
4.2.1.	Detekcja w oparciu o barwę	52
4.2.2.	Detekcja w oparciu o gradient	53
4.2.3.	Analiza predykcji metod składowych	53
4.3.	Badania zespołu metod	53
4.3.1.	Porównanie do istniejących metod	54
4.3.2.	Podsumowanie	55
Zakończenie		56
Bibliografia		57

SPIS RYSUNKÓW

1.1.	Paleta typu EPAL; a) zdjęcie rzeczywiste [źródło: clicktrans.pl]; b) wymiary [źródło: opel.fijalkowski.pl]	11
1.2.	Regał z paletami typu EPAL [źródło: rebstorage.com]	12
1.3.	Paleta typu EPAL z ładunkiem, owinięta folią typu "stretch" [źródło: gema.biz.pl]	12
1.4.	Wybrane typy palet; a) paleta typu EPAL drewniana; b) paleta typu EPAL bejcowana; c) paleta kartonowa; d) paleta z tloczonego drewna; e) paleta plastikowa zamknięta od dołu; f) paleta plastikowa otwarta od dołu; g) paleta metalowa; h) paleta plastikowa z możliwością zagnieździania [źródła: mygreenchapter.com, trust88.net, repalletize.com]	13
1.5.	Zdjęcia palet w różnych warunkach oświetleniowych; a) prześwietlone [źródło: [44]]; b) pora zmierzchu, brak oświetlenia sztucznego; c) równomierne, sztuczne oświetlenie	14
1.6.	Zdjęcia palet wykonane pod różnym kątem względem płaszczyzny podłożu; a) równolegle; b) pod kątem 10 stopni - symulacja przechylenia robota pod wpływem nierówności podłożu	14
1.7.	Zdjęcia palet przy mocowaniu kamery na różnej wysokości; a) kamera na wysokości połowy frontu palety; b) kamera na wysokości wielokrotnie przewyższającej wysokość frontu palety	14
1.8.	Zdjęcia palet wykonane pod różnym kątem względem płaszczyzny płaszczyzny frontu palety; a) równolegle, na wprost; b) pod kątem, z lewej strony; c) pod kątem, z prawej strony	15
2.1.	Etapy detekcji metodą Byun et al.[7]; a) obraz wejściowy; b) rezultat binaryzacji; c) rezultat detekcji krawędzi; d) dominujące krawędzie; e) linie dopasowane do krawędzi [źródło: Byun et al.[7]]	17
2.2.	Etapy detekcji metodą Chen et al.[9]; a) obraz wejściowy; b) rezultat binaryzacji; c) rezultat detekcji narożników [źródło: Chen et al. [9]]	17
2.3.	Etapy detekcji metodą Syu et al.[40]; a) wyznaczone obszary zainteresowania; b) wybrany obszar zainteresowania; c) rozkład szarości w procentach z zaznaczonym załamaniem dla wybranego obszaru zainteresowania; d) rezultat binaryzacji [źródło: Syu et al.[40]]	18
2.4.	Etapy detekcji metodą Pages et al.[29]; a) obraz wejściowy; b) rezultat binaryzacji; c) największy klaster; d) rezultat zastosowania operacji morfologicznych, e) rezultat wykrywania krawędzi i szkieletyzacji; f) rezultat wybrania pikseli o najmniejszej współrzędnej Y; g) piksele należące do bocznej krawędzi palety wyznaczonej przez transformator Hough'a [źródło: Pages et al.[29]]	19
2.5.	Etapy detekcji metodą Kim et al.[26]; a) obraz wejściowy w skali szarości; b) rezultat wykrywania krawędzi oraz linii prostych; c) rezultat łączenia równoległych linii w prostokąty; d) rezultat walidacji par prostokątów [źródło: Kim et al.[26]]	20
2.6.	Wybrane etapy detekcji metodą Cuciara et al.[12]; a) obraz wejściowy w skali szarości z wyznaczonym obszarem zainteresowania; b) rezultat detekcji krawędzi; c) rezultat klasteryzacji; d) umiejscowienie detektorów krawędzi [źródło: Cuciara et al.[12]]	21
2.7.	Dwa wybrane obrazy wynikowe detekcji metodą Varga et al. z roku 2014 [44]. Czerwonymi liniami zaznaczono granice obszaru detekcji. Zielonymi prostokątami zaznaczono wykryte palety. [źródło: Varga et al. 2014 [44]]	22
2.8.	Cztery z ośmiu kanałów ekstrachowanych w metodzie Varga et al. z roku 2015 [45]; a) obraz w skali szarości; b) wielkość gradientu; c) wielkość gradientu pod kątem 0 stopni; d) wielkość gradientu pod kątem 90 stopni [źródło: Varga et al. 2015 [45]]	22

2.9. Wizualizacja wybranych cech wprowadzonych w metodzie Varga et al. z roku 2016 [46]; a) siatka, na podstawie której wylicza się znormalizowaną różnicę par; b) obszary nóg, na podstawie których obliczana jest średnia wielkość gradientu [źródło: Varga et al. 2016 [46]]	23
2.10. Szum na obrazie wykonanym w niedoświetlonym magazynie przemysłowym; a) obraz oryginalny; b) obraz po wyrównaniu histogramu; c) obraz w skali szarości	27
2.11. Hierarchiczna struktura transformacyjnych metod odszumiania obrazów na podstawie publikacji przeglądowej Fan et al. [17]	29
2.12. Hierarchiczna struktura metod odszumiania obrazów w oparciu o sieci neuronowe na podstawie publikacji przeglądowej Fan et al. [17]	29
2.13. Hierarchiczna struktura klasycznych metod odszumiania obrazów na podstawie publikacji przeglądowej Fan et al. [17]	30
2.14. Porównanie filtrów alfa-beta oraz alfa-beta-gamma dla detekcji obiektu o ruchu jednostajnie przyspieszonym. Czarną linią oznaczono ruch jednostajnie przyspieszony, niebieską pomiar, czerwoną estymację filtra alfa-beta, a zieloną estymację filtra alfa-beta-gamma. [źródło: [36]]	33
 3.1. Architektura wysokiego poziomu badanej metody detekcji wraz z ewaluacją.	39
3.2. Wizualizacja pojedynczego okna wyznaczonego techniką okna przesuwnego; a) kolorem zielonym zaznaczono pojedyncze okno; b) obszar obrazu wycięty przez zaznaczone okno .	39
3.3. Wizualizacja kanałów koloru wykorzystywanych do klasyfikacji na przykładowym obrazie z paletą. Przykłady uwzględniają operację wyrównania histogramu.	41
3.4. Rezultat binaryzacji obrazu zawierającego paletę w kolorze jasnego drewna, z wykorzystaniem klasyfikatorów wytrenowanych na trzech kolorach; a) niebieskim; b) ciemnego drewna; c) jasnego drewna	41
3.5. Rezultat operacji morfologicznych; a) wejściowy obraz binarny; b) wynik operacji domknięcia; c) wynik operacji domknięcia i otwarcia	42
3.6. Okno przesuwne z zaznaczonymi obszarami wymaganymi do obliczenia funkcji dopasowania 3.1 oraz marginesami; w - szerokość okna, h - wysokość okna, zielony prostokąt - okno przesuwne; żółte wypełnienie - obszar palety; szare wypełnienie - obszar dziur	43
3.7. Potok przetwarzania metody detekcji w oparciu o barwę	43
3.8. Etapy przetwarzania obrazów w metodzie gradientowej; a) obraz wejściowy; b) obraz w skali szarości; c) obraz po nałożeniu filtra Gaussa; d) obraz po wyrównaniu histogramu; e) wielkość gradientu; od f) do k) zorientowana wielkość gradientu w sześciu zakresach kątów podanych w stopniach, gdzie zakresy to kolejno 0-60, 60-120, 120-180, 180-240, 240-300 oraz 300-360	45
3.9. Przykładowe wartości indeksu Jaccarda dla wybranych predykcji. Kolor niebieski to etykieta, kolor zielony to predykcja.	48
 4.1. Przykładowe klatki z ujęć ze zbioru danych. Dwa ostatnie zdjęcia są wynikiem operacji wyrównania histogramu, która ułatwia ludzkiemu oku dostrzeżenie palet zarejestrowanych w ciemnych warunkach oświetleniowych.	50
4.2. Przykładowe klatki z ujęć ze zbioru danych z Rys. 4.1 z naniesionymi etykietami.	51

SPIS TABEL

2.1.	Tabela z informacjami o zbiorze danych użytym do testowania przeanalizowanych metod.	24
2.2.	Tabela z informacjami o ewaluacji oraz wynikach przeanalizowanych metod. L - lokalizacja, D - detekcja	25
2.3.	Tabela z podsumowaniem przeanalizowanych metod detekcji.	26
2.4.	Tabela z informacją o zastosowanym algorytmie odszumiania dla każdej z przeanalizowanych metod detekcji.	31
2.5.	Tabela z informacją o zastosowanym filtrze wygładzania po czasie dla każdej z przeanalizowanych metod detekcji.	36
4.1.	Tabela z wynikami Naiwnego klasyfikatora bayesowskiego uczonego oraz testowanego na zbiorze pikseli należących do palet o danym kolorze oraz tła.	52
4.2.	Macierze pomyłek dla Naiwnego klasyfikatora bayesowskiego uczonego oraz testowanego na zbiorze pikseli należących do palet o danym kolorze oraz tła.	52
4.3.	Tabela z wynikami klasyfikatora Lasu losowego uczonego oraz testowanego na zbiorze prostokątów zawierających palety oraz tło.	53
4.4.	Macierz pomyłek dla klasyfikatora Lasu losowego uczonego oraz testowanego na zbiorze prostokątnych okien zawierających palety oraz tło.	53
4.5.	Tabela z wartościami średnich i odchyлеń standardowych prawdopodobieństw występowania palety w oknach przesuwnych w autorskim zbiorze danych wyznaczanych metodami opartymi na barwie i gradiencie.	54
4.6.	Tabela z wynikami ewaluacji metod bazowych i ich połączenia badanymi sposobami agregacji. .	54

STRESZCZENIE

Celem pracy było opracowanie oraz zbadanie metody detekcji palet wykorzystywanych w przemyśle w sekwencji wideo pochodzącej z autonomicznego wózka widłowego. Zapotrzebowanie na tego typu rozwiązanie wynika z dynamicznego rozwoju Przemysłu 4.0, który charakteryzuje się integracją sterowanych cyfrowo maszyn, m.in. robotów transportujących palety po powierzchni magazynowej.

Zaproponowana metoda jest rezultatem wnikliwej analizy dotychczasowych rozwiązań. Zauważono braki w trzech obszarach dotychczasowych badań. Najważniejszym i najbardziej obiecującym z nich jest brak integracji technik analizy obrazów w skali szarości z kolorowymi. Zaproponowano rozwiązanie oparte o obie techniki, polegające na pozyskiwaniu cech z obu typów obrazów. Opracowano także autorską metodę detekcji w oparciu o barwę, która niweluje wady dotychczasowych rozwiązań, a jednocześnie stanowi metodę składową rozwiązania zintegrowanego.

Do badań pozyskano zbiór sekwencji wideo symulujących najazd autonomicznego wózka widłowego z zamontowaną kamerą na paletę. W zbiorze znajduje się 29 scen, łącznie nieco ponad 2250 obrazów. W celu uczenia metod oznaczono ręcznie wszystkie palety występujące na obrazach poprzez zaznaczenie prostokątnym obszarem o określonych wymiarach.

Przeprowadzono badania na autorskim zbiorze danych z wykorzystaniem uniwersalnej miary indeks Jaccarda. Osiągnięto wynik wyższy niż wiodące metody. Główne ograniczenia metody to praca wyłącznie z kolorami palet występującymi w zbiorze treningowym oraz nieco większy nakład obliczeniowy w stosunku do metod pozabawionych analizy barwy. W celu ułatwienia zrozumienia problemu praca jest bogato ilustrowana.

ABSTRACT

The main goal of this thesis was development and test of industry pallet detection method on video sequence from autonomous forklift. Demand on this kind of solutions comes from dynamic progress of Industry 4.0, which characterizes with integration of digitally guided machines like robots transporting pallets on the warehouse surface.

Proposed method is result of profound existing solutions analysis. Three main shortcomings were noted. The most important and promising of them is lack of integration of grayscale and color image analysis techniques. Proposed solution utilizes both techniques, which extract features from both image types. Novel color based detection method has been developed, which removes disadvantages of existing solutions and acts as component of proposed integrated solution.

Dataset containing video sequences simulating ride of autonomous forklift with camera attached in front of pallet has been collected. Dataset consists of 29 scenes, above 2250 images totally. For detector learning purposes, all pallets have been marked by hand with rectangle of known size.

Widely used and universal Jaccard index was used to test developed methods on collected

dataset. Reached score was higher than score of leading-edge methods. The main limitations of solution are work with pallet colors present in dataset only and a bit higher computational complexity than for methods without color information analysis. Thesis is reachly illuminated to increase understandability of solved problem.

WSTĘP

Leonardo Da Vinci był pierwszą znaną postacią, która tworzyła projekty pojazdów autonomicznych. Inżynier wyprzedzał jednak znacznie swoją epokę. Pierwsza opublikowana próba stworzenia pojazdu bez kierowcy miała miejsce w Stanach Zjednoczonych w latach 20 ubiegłego wieku. Konstruktorzy sterowali zdalnie wozem z drugiego, jadącego za nim pojazdu. W latach 30-70 podejmowano próby kontroli pojazdu z wykorzystaniem specjalnie przygotowanej infrastruktury. Tworzono drogi z ukrytymi magnetycznymi paskami lub kablami sygnałowymi. Jednak dopiero w latach 80 rozpowszechniona została próba prof. Ernst Dickmanns'a z uniwersytetu w Monachium, która bazowała na komputerowym przetwarzaniu obrazów. Badacz zamontował kilka kamer na aucie marki Mercedes. Komputer znajdujący się na pokładzie wykonywał nowatorski algorytm symulujący ruch sakkadowe(mimowolne) ludzkiego oka. Pojazd VaMoRs poruszał się wyłącznie po drogach wyłączonych z użytku. Niedługo po tej próbie, na świecie powstały innowacyjne projekty dla samochodów osobowych, pojazdów wojskowych i wahadłowców kosmicznych. Obecnie zastosowanie komputerowego przetwarzania obrazów stało się powszechnne. Obszarem zastosowań szczególnie bliskim niniejszej pracy jest logistyka. Wraz z popularyzacją pojęcia Przemysłu 4.0 wzrasta liczba zintegrowanych autonomicznych pojazdów poruszających się po powierzchniach magazynowych, w szczególności wózków widłowych.

OPIS PROBLEMU

Robot wyposażony w widły do transportu palet ma za zadanie wjazd pod paletę, uniesienie jej, transport w wyznaczone miejsce, opuszczenie, wyjechanie widłami spod palety oraz przemieszczenie do kolejnego miejsca. Lokalizacja wybranych palet znana jest najczęściej dzięki systemowi GPS pracującemu na magazynie. Precyzyjny wjazd pod paletę wymaga analizy otoczenia przy pomocy urządzeń zamontowanych na wózku oraz detekcji. Najczęściej wykorzystywane urządzenia to lasery i kamery. Zastosowanie kamer jest rozwiązaniem dużo tańszym i właśnie takie przyjęto w ramach niniejszej pracy. Detekcja polega na przetworzeniu sekwencji obrazów w przestrzeni kolorów RGB i wskazaniu obszarów na każdym obrazie, które zawierają palety. Kolejnym problemem, który nie jest jednak adresowany w pracy, jest lokalizacja, czyli transformacja współrzędnych palety na obrazie na współrzędne trójwymiarowe w przestrzeni magazynu.

CEL PRACY

Celem pracy jest zapoznanie się z dotychczasowymi rozwiązaniami opisanego problemu oraz opracowanie metody, która pozwoliłaby wypełnić znalezione braki, przez co poprawić skuteczność detekcji. Zaproponowana metoda powinna integrować poznane rozwiązania i redukować ich wady. Ponadto wymagane jest zebranie własnego zbioru danych, gdyż żaden reprezentatywny zbiór nie został opublikowany. Należy także przeprowadzić badania, które pozwolą na porównanie zaproponowanej metody do wiodących rozwiązań. Testy powinny być wykonywane na tym samym zbiorze i z wykorzystaniem popularnych w dziedzinie miar. Celem pośrednim jest

udokumentowanie zaprojektowanego rozwiązania, przeprowadzonych badań oraz osiągniętych wyników, tak aby możliwe było ich powtórzenie bądź uwzględnienie w przyszłych badaniach.

STRUKTURA PRACY

Praca składa się z czterech rozdziałów. W rozdziale 1 znajduje się bogato ilustrowany opis poruszanego problemu. Na opis składa się umiejscowienie w szerszej rzeczywistości oraz zgłębienie szczegółów związane z detekcją palet. W rozdziale 2 umieszczono przegląd literaturowy. Zapoznano się zarówno z pracami podejmującymi bezpośrednio tematykę detekcji palet wykorzystywanej w autonomicznych pojazdach, jak i publikacjami dotyczącymi dwóch aspektów, w których zauważono istotne braki. Są to odszumianie obrazów i wygładzanie detekcji w sekwencji wideo. Przegląd literaturowy kończy się licznymi tabelami mającymi za zadanie ustrukturyzowanie pozyskanej wiedzy oraz wnioskami, na podstawie których wyznaczony został kierunek badań w dalszej części pracy. W rozdziale 3 znajduje się pełny opis zaproponowanej metody. Rozpoczyna się od omówienia ogólnej architektury, gdyż opracowane rozwiązanie składa się z dwóch metod składowych, a ponadto wykorzystuje technikę okna przesuwnego. Każdy aspekt został następnie szczegółowo opisany i wzbogacony o przykładowe ilustracje kolejnych etapów przetwarzania. Ostatni rozdział 4 zawiera opis pozyskanego zbioru danych, przeprowadzonych badań, otrzymanych wyników oraz podsumowanie. Praca kończy się opisowym zakończeniem oraz bibliografią.

1. OPIS PROBLEMU

1.1. UMIEJSCOWIENIE W SZERSZEJ RZECZYWISTOŚCI

W celu umiejscowienia problematyki podejmowanej przez niniejszą pracę w dzisiejszym świecie, warto odwołać się do zyskującego ostatnimi czasy popularność pojęcia "Przemysłu 4.0". Przemysł 4.0 to zbiorcze zagadnienie oznaczające integrację inteligentnych maszyn, systemów oraz wprowadzanie zmian w procesach produkcyjnych w celu zwiększenia wydajności wytwarzania. Jednym z najważniejszych obszarów tego szerokiego pojęcia jest automatyzacja. Zauważa się trend do automatyzowania wszystkich etapów produkcji występujących w przemyśle. Automatyzację można podzielić na dwie główne kategorie, czyli informatyzację oraz robotyzację, gdzie ta druga ma za zadanie usprawnienie procesów bardziej fizycznych. Procesem wspólnym dla niemal wszystkich dziedzin przemysłu czy handlu jest magazynowanie. Z kolei czynnością wspólną dla większości magazynów jest transport palet po powierzchni magazynowej. Tradycyjnie odbywa się to z wykorzystaniem wózków widłowych, którymi operują wykwalifikowani pracownicy. W dzisiejszych czasach dokonywane są jednak liczne próby zastąpienia operatora systemem informatycznym, który potrafiłby pokierować wózkiem widłowym. Wózek widłowy sterowany przez taki system nazywany jest często bardziej ogólnie robotem AGV (ang. Automated Guided Vehicle). Robot AGV służący do pracy z paletami składa się z na ogół z kilku podstawowych modułów. Są to moduł sterowania, bezpieczeństwa czy modelowania przestrzeni, a wśród nich moduł wizyjny, którego zadaniem jest detekcja palet, w celu następnego uniesienia oraz transportu. Autonomiczny wózek widłowy pracuje często w parze z systemem nawigacji znajdującym się w magazynie, który zna miejsca stacjonowania poszczególnych palet, punktów załadunku czy innych lokalizacji, w których powinna znaleźć się w danym momencie paleta takich jak np. maszyna do foliowania. System nawigacji obarczony jest jednak błędem, który sprawia, że precyzyjny wjazd wideł wózka pod paletę nie jest możliwy bez osobnego systemu wizyjnego zamontowanego na wózku.

Typowy scenariusz, który powinien zostać zrealizowany przez robota AGV pracującego z paletami składa się z kilku następujących etapów:

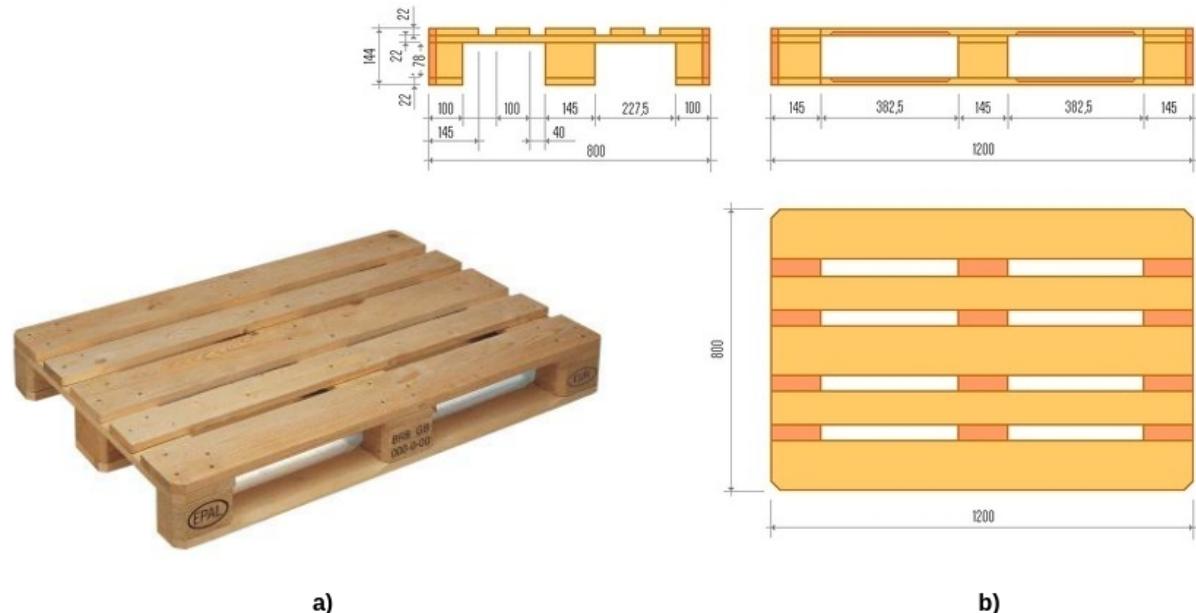
1. Otrzymanie lokalizacji palety z magazynowego systemu nawigacji
2. Przejechanie po magazynie do miejsca, w którym stoi paleta
3. Przełączenie na lokalizację na podstawie systemu wizyjnego na wózku
4. Detekcja oraz precyzyjna lokalizacja palety z wykorzystaniem systemu wizyjnego
5. Wjazd widłami pod paletę, z jednoczesną ciągłą detekcją oraz lokalizacją w celu korygowania pozycji
6. Uniesienie wideł
7. Przejazd w kolejne miejsce wyznaczone przez magazynowy system nawigacji

Realizacja całego scenariusza wymaga współpracy wielu systemów oraz algorytmów. W niniejszej pracy podjęty został problem detekcji palety na obrazie pochodzący z systemu wizyjnego.

1.2. SZCZEGÓŁY PROBLEMU DETEKCIJY PALET

Detekcja palet jest szczególnym przypadkiem zagadnienia detekcji obiektów na obrazie. Rozpoczynając rozpracowywanie problemu warto skupić się na cechach charakterystycznych wykrywanego obiektu oraz na właściwościach całego obrazu.

1.2.1. Cechy wykrywanego obiektu



Rys. 1.1. Paleta typu EPAL; a) zdjęcie rzeczywiste [źródło: clicktrans.pl]; b) wymiary [źródło: opel.fjalkowski.pl]

Na rysunku 1.1 możemy zauważyć, że paleta typu EPAL posiada wiele charakterystycznych kształtów w postaci "nóg" oraz odstępów pomiędzy nimi, które potocznie można nazwać "dziurami". Wymiary nóg, dziur oraz płaskich belek tworzących górną część palety są w tym przypadku ściśle określone. Należy zadać sobie pytanie o to, które z tych elementów mogą posłużyć do detekcji. Na rysunku 1.2, który przedstawia przykładowy regał z paletami, możemy zauważyć, że zarówno górną część palety jak i dłuższe jej boki są zasłonięte. Góra palety przykryta jest towarem, natomiast boki zakryte są przez inne palety, bądź pozostałe obiekty magazynowe jak np. słup należący do regału. Jedynie frontalna część palety pozostaje widoczna w każdym przypadku. Zauważa się więc istotne ograniczenie, polegające na tym, że detekcja całej palety na obrazie musi sprowadzić się do wykrywania samego tylko frontu, ponieważ w większości przypadków wyłącznie on jest widoczny. Ponadto do wejchania widłami nie jest potrzebna widoczność całego frontu, a jedynie samych dziur. Może bowiem zdarzyć się, że jego fragment również zostanie zakryty. Najczęściej zdarza się to przez owijanie towaru folią typu "stretch", której zadaniem jest zabezpieczenie ładunku przed spadaniem. Folia zakrywa wtedy górną część frontu palety utrudniając detekcję. Taką sytuację można zaobserwować na rysunku 1.3.

Zarówno kształty jak i barwy palet mogą się znaczco różnić. Najprostszym przypadkiem jest przyjęcie pewnego standardu np. EPAL i stosowanie go na całym magazynie. Jednak nie zawsze można przyjąć takie założenie w rzeczywistych rozwiązaniach. Jest to kwestia ściśle zależna od potrzeb i warunków klienta, który pragnie wykorzystywać roboty AGV. Kolejnym czynnikiem, który różni palety jest barwa, która może zmieniać się oczywiście również w obrębie tego samego



Rys. 1.2. Regał z paletami typu EPAL [źródło: rebstorage.com]



Rys. 1.3. Paleta typu EPAL z ładunkiem, owinięta folią typu "stretch" [źródło: gema.biz.pl]

kształtu. Barwa może wynikać z materiału, z którego wykonana jest paleta, naniesienia farby lub nawet dużego poziomu zabrudzenia, któremu z czasem ulegają palety drewniane eksploatowane poza suchymi i czystymi magazynami. Przykładowe kształty i kolory palet przedstawiono na rysunku 1.4. Najbardziej charakterystyczny element frontu palety, czyli nogi, różni się znaczco w pomiędzy przedstawionymi typami. W typie EPAL nogi są prostopadłościanami, w paletach plastikowych nogi są bardziej obły i stożkowane, natomiast w palecie , metalowej nogi są znacznie węższe. Drugi charakterystyczny element, czyli dziury również różni się kształtem, a ponadto mogą one być domknięte belką poprzeczną wyłącznie od góry jak w typie f) lub również od dołu jak w e). Domknięcie dziury wyłącznie od góry sprawia, że stajemy się bardziej zależni od barwy i odcienia podłoża na etapie detekcji, ponieważ dolną granicę dziur stanowi wtedy podłoga.

1.2.2. Wpływ warunków zewnętrznych

Kolejnym aspektem, który należy uwzględnić w problemie detekcji obiektu na obrazie są warunki w jakich pozyskiwany jest obraz zawierający obiekt. Mają one bowiem często kluczowy wpływ na skuteczność detekcji. Podobnie jak typ wykrywanej palety, tak samo warunki pracy robota AGV są w pełni zależne od klienta, który nabywa taką technologię. W dzisiejszym świecie drogie i zaawansowane technologicznie roboty pracują jednak w bardzo przyjaznych warunkach takich jak hale magazynowe, oświetlone, zamknięte, o równej, gładkiej powierzchni podłogi, z wy-



Rys. 1.4. Wybrane typy palet; a) paleta typu EPAL drewniana; b) paleta typu EPAL bejcowana; c) paleta kartonowa; d) paleta z tłoczonego drewna; e) paleta plastikowa zamknięta od dołu; f) paleta plastikowa otwarta od dołu; g) paleta metalowa; h) paleta plastikowa z możliwością zagnieźdzania [źródła: mygreenchapter.com, trust88.net, repalletize.com]

starcząca ilością miejsca do poruszania się pojazdu. Spośród wymienionych czynników, jednym z najbardziej istotnych jest oświetlenie. W skrajnym przypadku, brak oświetlenia uniemożliwia oczywiście detekcję na podstawie obrazu. Z drugiej strony, nadmierne oświetlenie pochodzące na przykład ze światła słonecznego, które skupia się przechodząc przez szyby, a następnie odbija od błyszczącej, równej posadzki powodując efekt przyćmienia lub przejaśnienia fragmentów palety, również może zaszkodzić skuteczności detekcji. Przykłady zdjęć palet pozyskanych w różnych warunkach oświetleniowych przedstawiono na rysunku 1.5. Innym ważnym czynnikiem jest podłożo, po którym porusza się robot, ponieważ ma ono bezpośredni wpływ na kąt pod jakim obserwowana jest paleta i jego zmiany. Jak nadmieniono wyżej, najczęściej jest to jednak równa i płaska posadzka magazynowa, która sprawia, że nie trzeba uwzględniać przechylania palety względem dolnej czy górnej krawędzi obrazu z kamery. Należy jednak pamiętać, że jeśli na posadzce mogą leżeć płaskie przedmioty takie jak fragmenty kartonów lub występują inne nierówności, na które może wjechać robot, to uwzględnienie odchyleń może być konieczne, a ewentualne założenie o równoległości krawędzi palety względem krawędzi obrazu obniży skuteczności detekcji. Przykład opisanej sytuacji przedstawiono na rysunku 1.6.

1.2.3. Pozycja kamery

Opisana kwestia nierówności podłożu jest związana ściśle z kwestią mocowania kamery na robocie, która również może mieć niemałe znaczenie. Przy założeniu, że robot stoi na przeciwko frontu palety, zamontowanie oka kamery zbyt wysoko sprawi, że front palety będzie widziany z góry i zostanie istotnie zwężony przez perspektywę. Z kolei mocowanie kamery zbyt nisko może doprowadzić do podobnego efektu, jednak o mniejszej skali, ponieważ podłoga stanowić będzie zawsze fizyczne ograniczenie. Optymalną wysokością wydaje się być umiejscowienie oka kamery w połowie wysokości frontu palety, dzięki czemu zaobserwujemy jego rzut o największej



Rys. 1.5. Zdjęcia palet w różnych warunkach oświetleniowych; a) prześwietlone [źródło: [44]]; b) pora zmierzchu, brak oświetlenia sztucznego; c) równomierne, sztuczne oświetlenie



Rys. 1.6. Zdjęcia palet wykonane pod różnym kątem względem płaszczyzny podłożu; a) równolegle; b) pod kątem 10 stopni - symulacja przechylenia robota pod wpływem nierówności podłożu

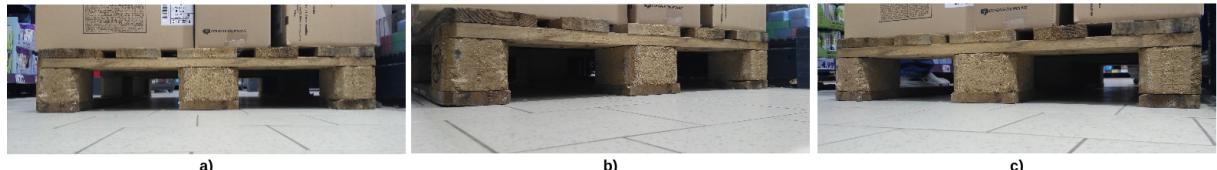
powierzchni oraz zminimalizujemy odkształcenie. W przypadku, gdy zakładamy obecność palet o kilku różnych wysokościach, jedną z dróg postępowania jest ich uśrednienie. Ujęcia z kamery zamocowanej na różnych wysokościach przedstawiono na rysunku 1.7.



Rys. 1.7. Zdjęcia palet przy mocowaniu kamery na różnej wysokości; a) kamera na wysokości połowy frontu palety; b) kamera na wysokości wielokrotnie przewyższającej wysokość frontu palety

Kolejną kwestią jest pozycja robota z kamerą względem palety, która determinuje kąt pod jakim jest ona obserwowana w trakcie wjeżdżania widłami, przy założeniu, że wysokość moco-

wania kamery na robocie jest znana i stała. Najprostszą sytuacją jest założenie, że zawsze pozycją wyjściową jest widok na wprost palety, równolegle do płaszczyzny frontu. Jest to założenie możliwe dzięki obecności magazynowego systemu nawigacyjnego, który z pewną niedokładnością przekazuje robotowi lokalizację palety. Może się jednak okazać, że niedokładność ta będzie zbyt duża lub alejka jest zbyt mała, żeby ustawić się równolegle do frontu palety. Wtedy wymagane będzie uwzględnienie perspektywicznego zniekształcenia, które sprawi, że jeden bok frontu palety będzie na obrazie wyższy od drugiego. Taką sytuację przedstawiono na rysunku 1.8.



Rys. 1.8. Zdjęcia palet wykonane pod różnym kątem względem płaszczyzny płaszczyzny frontu palety; a) równolegle, na wprost; b) pod kątem, z lewej strony; c) pod kątem, z prawej strony

Po zgłębieniu problemu detekcji palet na obrazie z kamery zamocowanej na autonomicznym wózku widłowym, zapoznano się z dostępnymi rozwiązaniami w rozdziale 2.

2. PRZEGŁĄD LITERATUROWY

Punktem wyjścia do przeprowadzenia badań jest zapoznanie się z istniejącymi publikacjami w celu poznania dotychczas dokonanych przez świat nauki postępów pod kątem rozwiązywanego problemu detekcji palet. Zapoznano się zarówno z kompletnymi rozwiązaniami związanymi bezpośrednio z problemem dotyczącym niniejszej pracy oraz tymi, które czerpią z dorobku całej dziedziny przetwarzania obrazu i wideo, dzięki czemu mogą posłużyć jako składowe etapy w procesie detekcji.

2.1. ROZWIĄZANIA DOTYCZĄCE BEZPOŚREDNIO DETEKCJI PALET

2.1.1. Wprowadzenie

Przegląd literaturowy rozpoczęto od odszukania i zrozumienia publikacji bezpośrednio rozwiązujących zaadresowany problem detekcji palet. Jest to krok kluczowy dla rozpoznania co zostało już osiągnięte i przebadane, dzięki czemu można będzie wskazać możliwe kierunki dla dalszych badań w tej pracy. Znaleziono 15 metod detekcji palet w środowisku wózka widłowego dokonujących detekcji na podstawie obrazów 2D z kamer. Inną istotną dla problemu odnogą rozwiązań jest detekcja na podstawie informacji o głębi obrazu pozyskowanej ze specjalnych sensorów. Jednak ze względu na charakter pracy wprost nawiązujący do dorobku dziedziny przetwarzania obrazów, publikacje dotyczące map punktów i informacji o głębi nie zostały przytoczone. Detekcja w oparciu o sam obraz rysuje się jako trudniejsze, ciekawsze i technicznie rzecz biorąc tańsze rozwiązanie.

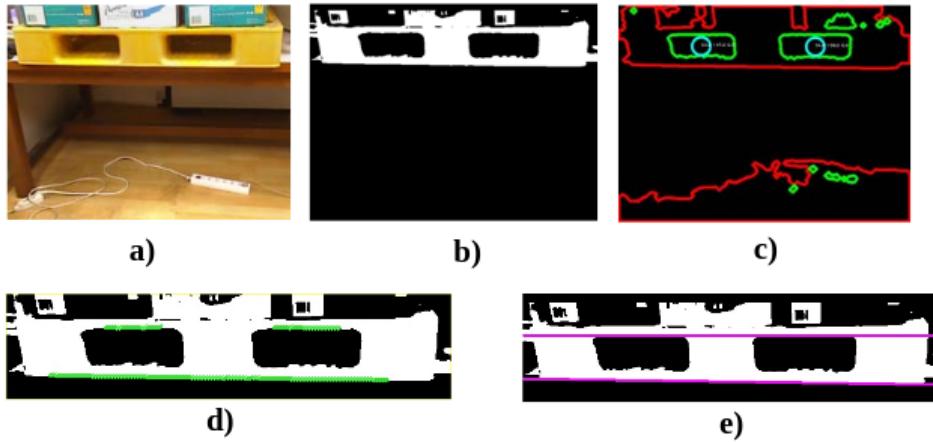
Dotychczasowe rozwiązania można podzielić na dwie główne rodziny. Pierwsza głównie wykorzystuje cechy związane z kolorem, nasyceniem i kontrastem do binaryzacji, natomiast druga opiera się na ekstrakcji cech z obrazów w skali szarości w postaci krawędzi, narożników czy wielkości gradientu. Ponadto znaleziono 2 publikacje Seelinger et al. [37] oraz Aref et al. [3], które dokonywały detekcji palety na podstawie znaczników znajdujących się na froncie. Jednak rozwiązania te sprowadzają się do dobrze zbadanego problemu odczytu kodów QR oraz pokrewnych, z tego powodu nie zostały omówione.

2.1.2. Metody oparte na binaryzacji

Metody Byun et al. [7] oraz Chen et al.[9] zakładają uproszczony przypadek plastikowych palet o jaskrawych kolorach (żółtym i czerwonym) oraz jednolitej fakturze, które mają równomiernie zaciennione dziury.

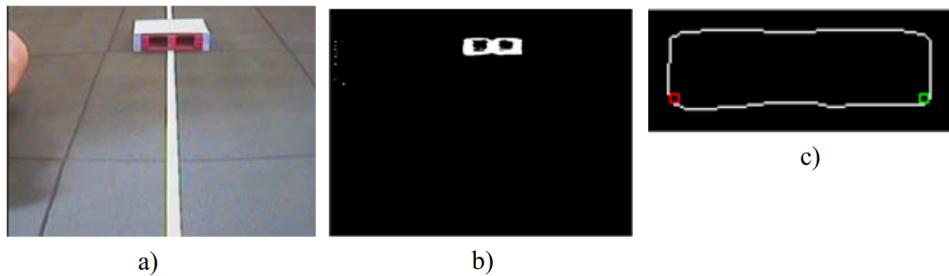
Pierwszym etapem **metody Byun et al.[7]** jest nieokreślona binaryzacja odróżniająca paletę od tła, po której stosowany jest filtr morfologiczny redukujący szum. Następnie wykonywana jest detekcja krawędzi, jednak wykorzystany algorytm nie jest podany. Kolejnym krokiem jest detekcja dwóch linii wyznaczających góre i dół palety. Dokonuje się ona poprzez dopasowanie dwóch linii do wykrytych krawędzi metodą najmniejszych kwadratów. Ostatnim etapem jest detekcja dziur palety pomiędzy dwiema najdłuższymi liniami, na podstawie wykrytych wcześniej

krawędzi oraz znanego kształtu i wymiarów tych dziur. Jak we wcześniejszych krokach, metoda dopasowania krawędzi do dziur palety nie jest jednak dokładnie opisana.



Rys. 2.1. Etapy detekcji metodą Byun et al.[7]; a) obraz wejściowy; b) rezultat binaryzacji; c) rezultat detekcji krawędzi; d) dominujące krawędzie; e) linie dopasowane do krawędzi [źródło: Byun et al.[7]]

Metoda Chen et al.[9] również rozpoczyna się od binaryzacji odróżniającej paletę od tła. Jednak tym razem znamy sposób podziału. Dokonuje się on na podstawie kanałów HSV i YUV, poprzez przyjęcie stałych, ręcznie określonych progów wartości dla każdego kanału. Kolejnym krokiem jest wykonanie dylatacji oraz erozji na binarnym obrazie w cel zapełnienia dziur. Po tej operacji na obrazie binarnym oczekujemy obserwacji obszaru zbliżonego kształtem do prostokąta o wielkości frontu palety. Ostatnim krokiem jest detekcja narożników metodą SUSAN[38].



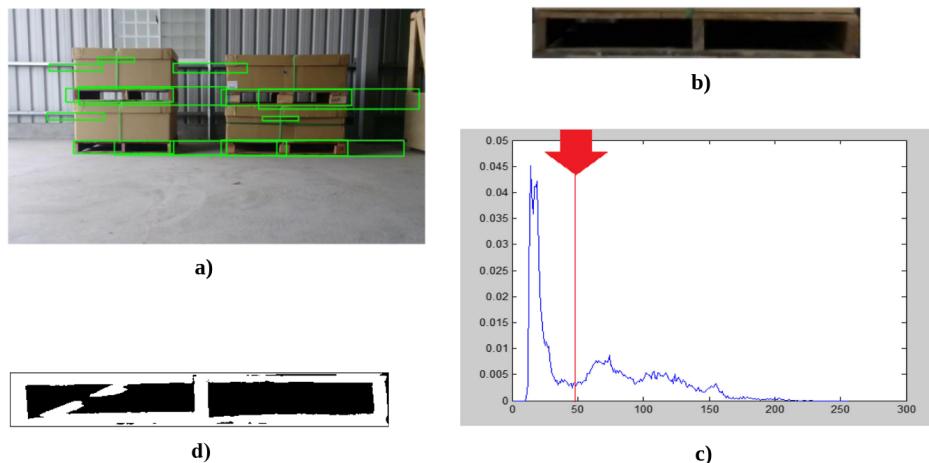
Rys. 2.2. Etapy detekcji metodą Chen et al.[9]; a) obraz wejściowy; b) rezultat binaryzacji; c) rezultat detekcji narożników [źródło: Chen et al. [9]]

Niestety dominujące ostatnio w magazynach drewniane palety, w przeciwieństwie do plastikowych, mogą mieć bardzo nierównomierną fakturę oraz odcień drewna i to w obrębie tej samej palety. To wyklucza powyższe proste metody, oparte na ręcznie określonych progach binaryzacji, z uniwersalnego zastosowania w przemyśle.

Metoda Syu et al.[40] podobnie jak dwie poprzednie zakłada, że zacienienie dziur pozwala na odróżnienie ich od frontu palety. Różnica jest natomiast taka, że została zbadana na drewnianych paletach, czego brakowało dwóm poprzednim metodom.

Metoda Syu et al.[40] rozpoczyna się od zmniejszenia rozdzielcości zdjęcia w celu szybszego przetwarzania metodą Symmetric Mask-based Discrete Wavelet Transform[8]. Kolejnym krokiem jest znalezienie obszarów zainteresowania algorytmem AdaBoost w połączeniu z klasyfikatorem kaskadowym z wykorzystaniem cech pozyskanych z prostokątów Haar'a. Jest to

rozwiążanie znane z popularnej metody detekcji twarzy Viola & Jones[48]. Rozwiążanie to jest zasadne ponieważ front palety ma kształt prostokąta, w którym znajdują się prostokątne ciemniejsze dziury. W taki kształt palety skutecznie wpasowują się prostokąty Haar'a, które wykrywają różnice w kontraste pomiędzy wybranymi obszarami. Kolejnym etapem jest binaryzacja w obszarze zainteresowania, polegająca na odróżnieniu frontu palety od dziur. Odbywa się ona na podstawie progu wskazanego przez załamanie w rozkładzie szarości. Ostatnim etapem jest obliczenie pokrywania się wyniku binaryzacji ze wzorcowym modelem palety. Ponadto opisana jest autorska metoda śledzenia palety na sekwencji obrazów oparta o nakładanie się obszarów wyznaczonych przez binaryzację w oparciu o histogram szarości.

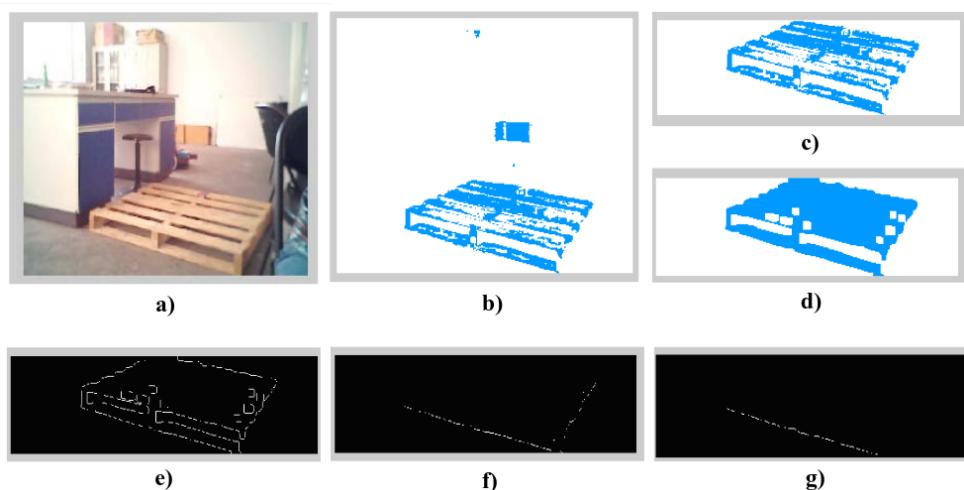


Rys. 2.3. Etapy detekcji metodą Syu et al.[40]; a) wyznaczone obszary zainteresowania; b) wybrany obszar zainteresowania; c) rozkład szarości w procentach z zaznaczonym załamaniem dla wybranego obszaru zainteresowania; d) rezultat binaryzacji [źródło: Syu et al.[40]]

W paletach drewnianych nie możemy jednak w rzeczywistości liczyć na zacienienie dziur, na którym mocno opierają się wszystkie powyższe metody. Zwłaszcza, gdy paleta stoi bez ładunku, a światło magazynowe prześwituje przez odstępy w deskach lub gdy ładunek jest obecny, lecz światło świeci od tyłu. Są to przypadki, których analizy nie podejmują autorzy powyższych publikacji. Utrudnienie polega na tym, że zmniejsza się wtedy różnica kontrastu pomiędzy dziurami, a frontem palety. To rzuca cień podejrzeń, że metody te nie są wystarczająco skuteczne w rzeczywistych warunkach magazynowych.

Metody Pages et al.[29] oraz Cui et al.[22] dokonują detekcji na podstawie widoku na paletę z góry i z boku. Posiadają niemal identyczny potok przetwarzania, a ich rezultatem jest wskazanie współrzędnych prostych wyznaczających dłuższy i krótszy bok palety. Zakładają, że kolor podłogi i ścian magazynu różni się wyraźnie od koloru palety. Potok przetwarzania rozpoczyna się od binaryzacji dokonywanej na podstawie stałych, przyjętych ręcznie wartości cech dla pojedynczych pikseli bez uwzględnienia ich sąsiedztwa. Kolejnym krokiem jest znalezienie największego klastra, czyli spodziewanej palety, w celu odrzucenia pozytywnie zaklasyfikowanych przedmiotów o barwie podobnej do palety. Następnie stosowane są kolejno operacja domknięcia i otwarcia w celu odszumienia. W kolejnym kroku wykrywane są krawędzie z wykorzystaniem filtra Sobel'a. Następnym krokiem jest szkieletyzacja wykrytych krawędzi. Algorytm użyty do szkieletyzacji nie jest niestety podany w żadnej z metod. W celu odrzucenia nieistotnych krawędzi i pozostawienia tylko tych, które należą do bocznej i frontowej części palety, w przedostatnim kroku usuwane są wszystkie piksele poza ostatnim w każdej kolumnie obrazu. Na koniec, z wykorzystaniem transformaty Hough'a, wyznaczane są współrzędne dwóch

linii należące do bocznej oraz frontowej krawędzi palety. Tym krokiem kończy się potok przetwarzania. Łatwo zauważać, że metody te nie mają niestety zastosowania w autonomicznych wózkach widłowych, w których zamocowanie kamery obserwującej paletę od boku, podczas gdy wózek wykonuje najazd od przodu, byłoby niewygodne i wymagało długiego ramienia, na którym byłaby umocowana kamera. Zastosowany potok przetwarzania może sprawdzić się jedynie w warunkach zaprezentowanych w obu publikacjach, czyli gdy kamera jest umieszczona gdzieś w pomieszczeniu, niezależnie od wózka. W praktyce paleta mogłaby być przesłonięta dowolnym obiektem występującym na magazynie takim jak inny wózek, inna paleta z towarem, regał czy przechodzący lub stojący pracownik. Prawdopodobnie z tego powodu wszystkie pozostałe prace zakładają, że kamera umocowana jest na wózku, a paleta obserwowana na wprost.

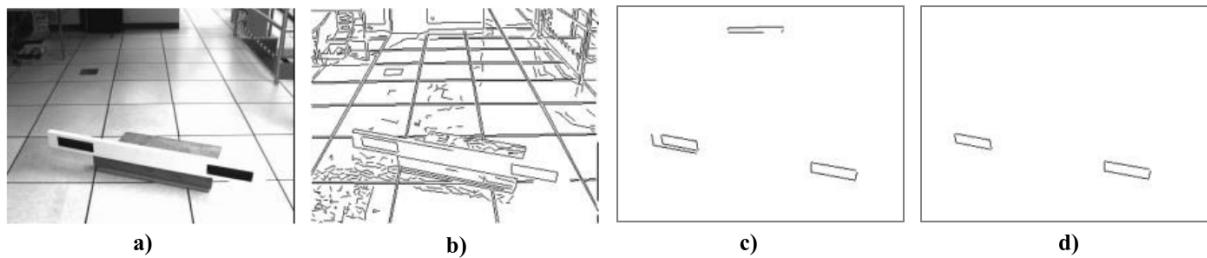


Rys. 2.4. Etapy detekcji metodą Pages et al.[29]; a) obraz wejściowy; b) rezultat binaryzacji; c) największy klaster; d) rezultat zastosowania operacji morfologicznych; e) rezultat wykrywania krawędzi i szkieletyzacji; f) rezultat wybrania pikseli o najmniejszej współrzędnej Y; g) piksele należące do bocznej krawędzi palety wyznaczone przez transformatę Hough'a [źródło: Pages et al.[29]]

2.1.3. Metody oparte na ekstrakcji cech z obrazów w skali szarości

Metody Kim et al.[26], Garibotto et al. z roku 1996[19] oraz z roku 1997[20], Cucciara et al.[12] oraz Varga et al. z 2014 roku [44], 2015 roku [45] i 2016 roku [46] stanowią odrębną klasę rozwiązań, ponieważ w przeciwieństwie do wszystkich wcześniej omówionych metod, bazują na ekstrakcji cech palety z obrazu w skali szarości bez uprzedniej binaryzacji.

Metoda Kim et al.[26], podobnie jak omówione wcześniej metody Byun et al.[7] oraz Chen et al.[9], zakłada uproszczony przypadek plastikowej palety. Rozpoczyna się od wykrycia krawędzi algorytmem Canny. Drugim krokiem jest wykrycie linii prostych detektorem Lowe'a. W trzecim kroku łączone są wykryte linie proste. W przedostatnim kroku, z wykorzystaniem warunków o równoległość, linie proste są łączone w prostokąty. Ostatnim etapem jest ocenienie każdej pary znalezionych prostokątów pod kątem dopasowania do wzorcowych wymiarów dziur palety oraz wybranie pary o najlepszym dopasowaniu. W oryginalnej pracy zamieszczono obrazy kolejnych kroków detekcji wyłącznie dla atrapy palety przedstawione na Rys. 2.9.



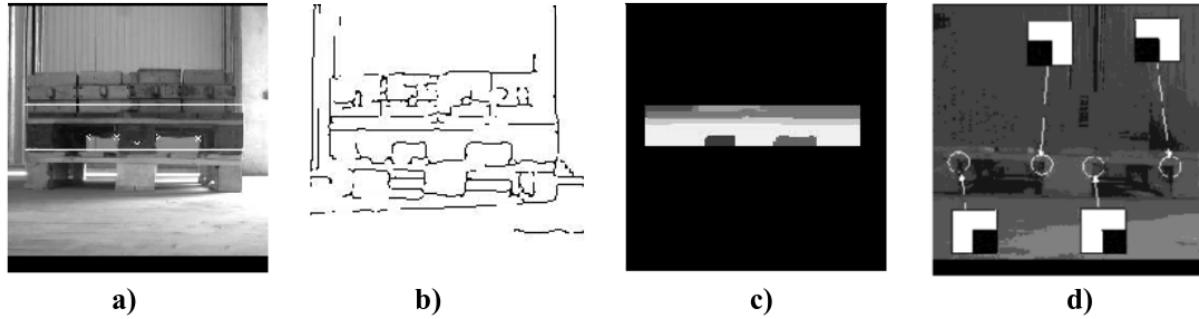
Rys. 2.5. Etapy detekcji metodą Kim et al.[26]; a) obraz wejściowy w skali szarości; b) rezultat wykrywania krawędzi oraz linii prostych; c) rezultat łączenia równoległych linii w prostokąty; d) rezultat walidacji par prostokątów [źródło: Kim et al.[26]]

Metody Garibotto et al.[20] [19] są opisane bardzo pobieżnie. Wykorzystują algorytm Region Growing do segmentacji według kontrastu. Spodziewanym wynikiem są 3 segmenty oznaczające dwie zacienione dziury i niezacieniony front plastikowej palety domknietej od dołu, o znanym kształcie.

Metody Kim et al.[26] oraz Garibotto et al.[19][20] te nie zostały niestety przebadane na drewnianych paletach, w których detekcja krawędzi może być mniej trywialnym problemem, ze względu na możliwość wpadania światła przez odstępy między deskami górnej płaszczyzny palety. W przeciwieństwie do nich, cztery kolejne metody zostały przebadane na paletach drewnianych, a ponadto w mocno zróżnicowanych warunkach oświetleniowych.

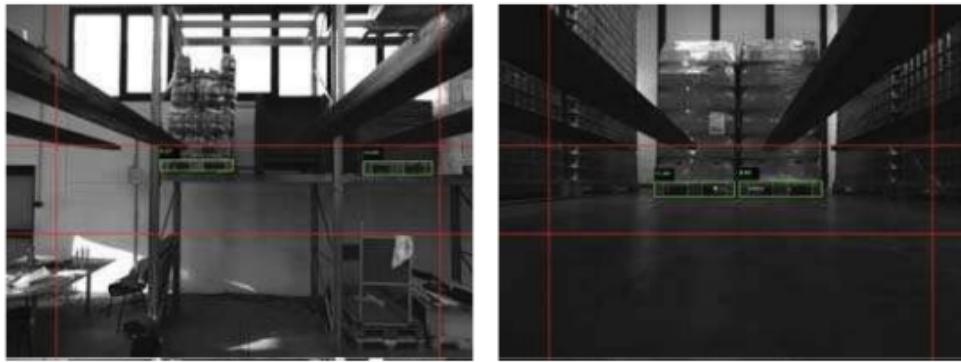
Metoda Cucciara et al.[12] rozpoczyna potok przetwarzania od operacji rozciągnięcia kontrastu. Operacja ta ma na celu unieważnienie metody na zmienne warunki oświetleniowe. Kolejnym krokiem jest wyznaczenie obszarów zainteresowania. Obszary te wskazują na miejsca na obrazie, w których najbardziej możemy spodziewać się palety. Wyznaczenie ich pozwala na odrzucenie pozostałych obszarów obrazu, w których najprawdopodobniej nie ma palety, a tym samym skrócenie czasu działania. Do wykrycia tych obszarów wykorzystywana jest Skorelowana Transformata Hough'a [11], która jest wzbogaconą wersją Gradientowej Transformaty Hough'a [28]. Różnica między dwiema transformatami polega na tym, że wersja skorelowana dokonuje detekcji dwóch blisko położonych, niemal równoległych linii na raz, w oparciu o gradient pikseli. Te dwie równoległe linie mogą bowiem wyznaczać płaszczyznę lądowania frontu palety. Wykrywanie obszarów zainteresowania przebiega od góry do dołu obrazu, a każda para znalezionych linii jest zapamiętywana, aby następnie w jej sąsiedztwie dokonać ekstrakcji cech oraz klasyfikacji. Proces ekstrakcji cech polega na znalezieniu wirtualnych oraz rzeczywistych narożników w obszarze zainteresowania. Oba zadania rozpoczynają się od detekcji krawędzi algorytmem Canny. Wirtualne narożniki stanowią przecięcie poziomej linii znalezionej na etapie wyszukiwania obszaru zainteresowania z pionowymi krawędziami wyznaczonymi przez zastosowanie Transformaty Hough'a dla krawędzi. Z kolei rzeczywiste narożniki wykrywane są w kilku krokach. Pierwszym krokiem jest klasteryzacja obszaru zainteresowania algorytmem Hybrid Linkage Region Growing [23] w oczekiwaniu, że wyodrębnii się w ten sposób osobno obszar obu dziur i frontu palety. Kolejnym krokiem jest wybranie heurystycznie dwóch największych klastrów, w przypadku, gdy zostało wykrytych więcej. Ostatnim krokiem jest wykrycie narożników pomiędzy nogami palety, a górną płaszczyzną poprzez przykładanie masek czterech oczekiwanych narożników do sklasteryzowanego obrazu i wybór najlepszych dopasowań jako znalezionych narożników. Metoda wykrywania narożników opisana jest szerzej w Trucco Verri[43]. Po wykryciu wirtualnych oraz rzeczywistych narożników następuje klasyfikacja. Wykonywana jest przez zespół klasyfikatorów będących dwoma drzewami decyzyjnymi C4.5, po jednym dla narożników rzeczywistych i wirtualnych. W zbiorze uczącym wyróżniono 7 klas w zależności od widoczności narożników

palety. Duża liczba klas pozwala na klasyfikację np. połowy palety. Łączenie wyników obu klasyfikatorów odbywa się tak, że priorytet ma klasyfikator narożników rzeczywistych. W przypadku, gdy nie wykryje on żadnej klasy palety, podejmowana jest decyzja na podstawie klasyfikatora narożników wirtualnych. Zespół wykazuje w badaniach wyższą skuteczność niż pojedyncze klasyfikatory.



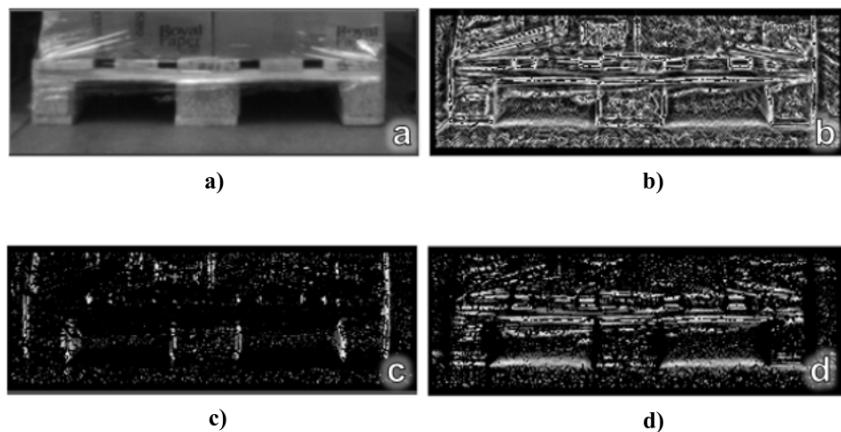
Rys. 2.6. Wybrane etapy detekcji metodą Cucciara et al.[12]; a) obraz wejściowy w skali szarości z wyznaczonym obszarem zainteresowania; b) rezultat detekcji krawędzi; c) rezultat klasteryzacji; d) umiejscowienie detektorów krawędzi [źródło: Cucciara et al.[12]]

Metoda Varga et al. z roku 2014 [44] rozpoczyna potok przetwarzania od konwersji do obrazu w skali szarości. Kolejnym krokiem przetwarzania wstępne jest zastosowanie filtra Gaussa, który wygładza krawędzie na obrazie oraz zamazuje detale. Autorzy publikacji kładą duży nacisk na duże znaczenie operacji wyrównania histogramu, która łagodzi wpływ niekorzystnych i niejednorodnych warunków oświetleniowych oraz zwiększa wartość gradientu. Kolejny krok dotyczy optymalizacji czasowej i polega na prostym odrzuceniu obszarów u góry, dołu oraz po bokach obrazu. Możemy tego dokonać dzięki wiedzy o około centralnym umiejscowieniu palety w każdym przypadku oraz estymacji maksymalnej wielkości palety na obrazie na podstawie znanej minimalnej odległości od oka kamery do frontu palety. Zasadnicze przetwarzanie rozpoczyna się od wykrycia krawędzi filtrem Sobel'a dla obu osi. Na podstawie wykrytych krawędzi wyznaczane są dominujące linie poziome na podstawie zakumulowanej wartości gradientu dla każdej poziomej linii obrazu. Liczba wykrywanych linii poziomych stanowi parametr metody. Następnie wykrywane są linie pionowe, które znajdują się pomiędzy dwiema danymi liniami poziomymi. Dalej znalezione linie są walidowane pod kątem dopasowania do modelu frontu palety. Spodziewamy się dwóch linii poziomych, pomiędzy którymi znajdzie się 6 linii pionowych wyznaczających krawędzie nóg palety. Pomyślna walidacja oznacza znalezienie obszaru zainteresowania. Na tym jednak nie kończy się detekcja, ponieważ dla jednego obrazu autorzy publikacji otrzymywali nawet do 300 tysięcy potencjalnych obszarów palet. W kolejnym kroku dla każdego obszaru zainteresowania liczone są 24 cechy związane ze średnią intensywnością, odchyleniem standardowym intensywności, siłą krawędzi oraz dysproporcją dla podobszarów wyznaczonych przez pionowe i poziome linie. Niektóre z cech to różnice między cechami wybranych, pojedynczych podobszarów. Po etapie ekstrakcji cech następuje klasyfikacja algorytmem AdaBoost z wykorzystaniem tysiąca dwupoziomowych drzew decyzyjnych. Za paletę uznawany jest obszar zainteresowania, który przekracza ustaloną wartość progu pewności klasyfikacji. Jednak to nie koniec, ponieważ dla danej rzeczywistej palety na obrazie możemy wykryć wiele zbliżonych obszarów powyżej progu. Ostatnim etapem jest wybór jednego obszaru dla każdej palety spośród wielu redundantnych, nakładających się obszarów na podstawie najwyższej wartości klasyfikacji.



Rys. 2.7. Dwa wybrane obrazy wynikowe detekcji metodą Varga et al. z roku 2014 [44]. Czerwonymi liniami zaznaczono granice obszaru detekcji. Zielonymi prostokątami zaznaczono wykryte palety. [źródło: Varga et al. 2014 [44]]

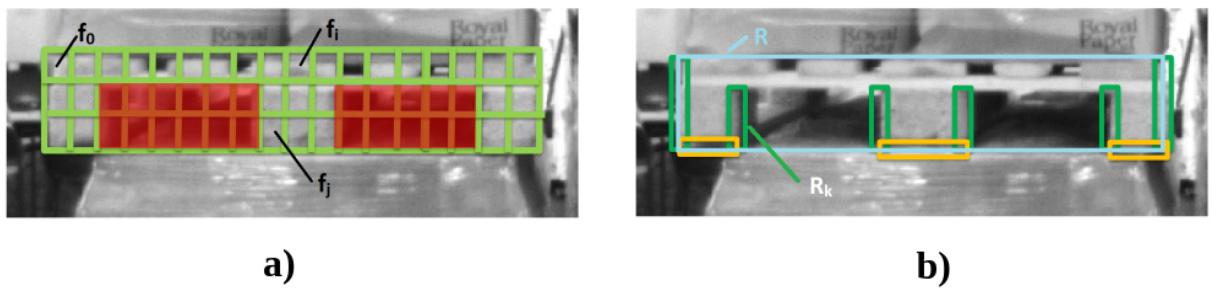
Metoda Varga et al. z roku 2015 [45] jest udoskonaleniem metody z roku wcześniej tych samych autorów. Wprowadza nowy sposób ekstrakcji cech naśladujący cechy Aggregate Channel Features [15]. Polega on na zastosowaniu ramki o stałym rozmiarze, którą przykładamy do obrazu w różnej skali. Z każdego przeskalowanego obrazu uzyskujemy 8 kanałów. Są to kolejno obraz w sali szarości, wielkość gradientu oraz zorientowana wielkość gradientu pod 6 różnymi kątami. Następnie wszystkie 8 kanałów agregujemy do jednego poprzez uśrednienie wartości. Wynikowy wektor 100 pikseli stanowi wejście do klasyfikatora będącego jak rok wcześniej algorytmem AdaBoost. Wprowadzono również zmiany w parametrach klasyfikatora. Liczbę drzew zwiększyliśmy z 1000 do 2048. Cały proces uczenia odbywa się zgodnie z protokołem 3 rund bootstrapingu opisany w [15]. Zastosowanie zaproponowanych zmian pozwoliło na wzrost precyzyji detekcji o 20 procent.



Rys. 2.8. Cztery z ośmiu kanałów ekstrahowanych w metodzie Varga et al. z roku 2015 [45]; a) obraz w skali szarości; b) wielkość gradientu; c) wielkość gradientu pod kątem 0 stopni; d) wielkość gradientu pod kątem 90 stopni [źródło: Varga et al. 2015 [45]]

Metoda Varga et al. z roku 2016 [46] również wprowadza zmiany w stosunku do metody z roku 2014 tych samych autorów. Pierwsza zmiana dotyczy sposobu wykrywania krawędzi. Zbadano do tego celu metodę znormalizowanych wartości gradientu zaproponowaną w publikacji z 2005 roku dotyczącej detekcji osób na obrazie [14]. Polega ona na podzieleniu wartości gradientu przez wartości uzyskane z prostego filtra konwolucyjnego. Taki zabieg ma na celu otrzymanie równomiernie skutecznej detekcji krawędzi zarówno w ciemnych jak i jasnych obszarach tego

samego obrazu. Na podstawie wartości znormalizowanego gradientu wyszukiwane są obszary zainteresowania tak samo jak w pierwotnej metodzie. Kolejnym krokiem jest ekstrakcja cech. Wprowadzono 3 rodzaje cech. Pierwszy rodzaj został nazwany znormalizowanymi różnicami par i polega na podzieleniu przesuwnego okna o stałym rozmiarze 5x20 na określona liczbę kwadratów. Każdy kwadrat wyznacza pewien niewielki obszar palety, który po zmniejszeniu rozdzielczości staje się pojedynczym pikselem. Różnice pomiędzy wartościami wszystkich pikseli stanowią 1350 cech. Obszary dziur w paletach nie są brane pod uwagę. Wartości wszystkich cech poddawane są dodatkowo funkcji sigmoidalnej. Kolejnym rodzajem cech są wartości gradientu w 9 obszarach, w których spodziewane są krawędzie nóg palety. Wartość gradientu dla każdego obszaru jest uśredniana. Trzecim rodzajem cech jest 256 wartości wynikających z zastosowania deskryptorów LBP (Local Binary Patterns) [24] na obszarze wyznaczonym przez okno przesuwne. W zależności od przyjętych hiperparametrów, wprowadzenie zmian spowodowało wzrost precyzji detekcji o wartość od 8 do 20 procent.



Rys. 2.9. Wizualizacja wybranych cech wprowadzonych w metodzie Varga et al. z roku 2016 [46]; a) siatka, na podstawie której wylicza się znormalizowaną różnicę par; b) obszary nóg, na podstawie których obliczana jest średnia wielkość gradientu [źródło: Varga et al. 2016 [46]]

2.1.4. Ewaluacja i wyniki

W tabeli 2.2 przedstawiono informacje o sposobie ewaluacji oraz wynikach każdej z metod. Natomiast w tabeli 2.1 przedstawiono informacje o zbiorze danych, na którym przeprowadzono badania. We wszystkich metodach poza Garibotto et al. [19][20] opisano sposób ewaluacji. Spośród 10 metod wyłącznie 5 uwzględnia ewaluację detekcji, nie tylko samej lokalizacji. Wśród tych 5 metod można wyróżnić 4 odmienne sposoby ewaluacji. Wyłącznie 4 z 5 metod zbadano na reprezentatywnej próbce tysięcy obrazów, są to Syu et al. [40] oraz prace Varga et al. z trzech kolejnych lat [44][45][46]. Wszystkie te prace mierzą procent wykrytych palet w stosunku do liczby wszystkich palet. Jest to niestety miara ignorująca przypadki, w których paleta została wykryta w miejscu, w którym jej nie ma. Z tego powodu wydaje się to być wysoce niemiardajny sposób walidacji. Jednak w pracy Varga et al. z roku 2014 [44] w sekcji "C. Detection postprocessing" w rozdziale "Proposed approach", możemy przeczytać, że liczba palet na obrazie znana jest z góry. Wynika to z faktu, że system wizyjny mocowany w autonomicznych wózkach widłowych służy do precyzyjnego określenia lokalizacji palety, której współrzędne są znane dzięki współpracującemu z nim systemowi nawigacji w magazynie. System nawigacji obarczony jest błędem uniemożliwiającym bezpieczne wejchanie wideł pod palety, jednak wystarcza do wskazania robotowi, w którym miejscu magazynu znajdują się palety. Przy takim założeniu ewaluacja uwzględniająca tylko poprawnie wykryte palety nabiera pełnego sensu. Istnieje jednak kolejny problem dotyczący ewaluacji, który obecny jest w każdej z 4 prac, a nie został rozwiązany.

Polega na tym, że fakt poprawnego wykrycia palety jest bardzo różnie definiowany. W pracy Varga et al. z roku 2014 [44] wprowadzono definicje przedstawioną we wzorze 2.1

$$\begin{aligned} |D.width - C.width| &< 10px \\ C.height/D.height &> 0.7 \end{aligned} \quad (2.1)$$

gdzie:

- A = pole wypredykowanego prostokąta
- B = pole prostokątnej etykiety
- C = pole przecięcia A i B
- D = pole sumy A i B
- $width$ = szerokość pola prostokąta
- $height$ = szerokość pola prostokąta

Z kolei w pracy tych samych autorów opublikowanej rok później przyjęto odmienną definicję:

$$\begin{aligned} D.width - C.width &\leq 15px \\ D.height - C.height &\leq 15px \end{aligned} \quad (2.2)$$

gdzie zmienne te same co na wykresie 2.1.

Natomiast pracy Syu et al. [40] poprawna detekcja określana jest na podstawie centralnego punktu frontu palety mieszczącego się w obrębie środkowej nogi, jednak dokładny wzór nie został podany. Łatwo więc zauważyc, że sposoby ewaluacji zaproponowane w przytoczonych pracach są nieporównywalne, a ponadto w pracach autorów Varga et al. są one zależne od ręcznie wyznaczonych wartości pikseli stanowiących progi. Wartości te są silnie uzależnione od rozdzielczości obrazów wykorzystanych do badań. Wprowadzenie innych rozdzielczości wymagałoby skalowania progów lub obrazów. Ponadto konkretne wartości takie jak 10px i 0.7 we wzorze 2.1 oraz 15px we wzorze 2.2 zostały podane przez autorów bez uzasadnienia, przez co nie przedstawiają żadnej wartości badawczej i nie ma powodu, by powielać je w kolejnych badaniach. W celu uniezależnienia się od ręcznie określanych progów zaproponowano metodę ewaluacji niewrażliwą na zmienną rozdzielczość obrazu, która bada precyzję zamiast liczby poprawnych detekcji. Autorską metodę opisano w rozdziale 3.6.

	Zbiór danych
Byun et al.	1 paleta, 100 obrazów
Chen et al	1 paleta, 70 obrazów
Syu et al.	10 scen, 1400 obrazów, 3000 oetykietowanych palet
Pages et al.	1 paleta, 50 obrazów, paleta w odległości do 3 m pod różnymi kątami
Cui et al.	1 paleta, 30 obrazów, paleta w odległości do 5 m pod różnymi kątami
Garibotto et al.	brak informacji
Kim et al.	1 paleta, 1 scena, 3 różne kąty obserwacji, 4 różne odległości od palety
Cucciara et al.	56 obrazów
Varga et al. 2014	37 scen, 6070 oetykietowanych palet
Varga et al. 2015	7124 obrazów, 9047 oetykietowanych palet
Varga et al. 2016	jak w Varga et al. z 2014 roku

Tabela 2.1. Tabela z informacjami o zbiorze danych użytym do testowania przeanalizowanych metod.

	Zadanie	Badana wartość	Miara	Wynik
Byun et al.	L	odchylenie standardowe pozycji x,y zwracanej przez system podczas ruchu kamery do przodu i do tyłu, nie wymaga zmierzenia wartości rzeczywistej	odchylenie standardowe	0.84 cm
Chen et al	L	błąd po uśrednieniu wartości z 10 klatek w tej samej pozycji, w odległości metra od palety	bezwzględna różnica od wartości rzeczywistej	1.6 cm bez filtracji, 2 cm z użyciem filtra Kalman'a
Syu et al.	D	błąd uśredniony po 10 scenach	procent niewykrytych palet (FN)	1.79 %
Pages et al.	L	błąd uśredniony po 50 klatkach dla dwóch skrajnych punktów palety widzianej od boku o współrzędnych x,y	średnia bezwzględna różnica od wartości rzeczywistej	0.315 cm na osi X, 0.125 cm na osi Y
Cui et al.	L	błąd uśredniony po 30 klatkach dla centralnego punktu palety widzianej od boku o współrzędnych x,y,z	średnia bezwzględna różnica od wartości rzeczywistej	0.106 cm na osi X, 0.065 cm na osi Y
Garibotto et al.	-	-	-	-
Kim et al.	L	błąd na nieznanej liczbie klatek w statycznej scenie dla centralnego punktu palety o współrzędnych x,y,z	odchylenie standardowe od wartości rzeczywistej	0.2 cm - średnia dla trzech osi
Cucciara et al.	D	liczba niewykrytych palet	suma	3
Varga et al. 2014	D	wykrycie i niewykrycie palety określane są jako zależność sumy i przecięcia obszarów etykiety z detekcją	procent wykrytych palet (TP)/procent niewykrytych palet(FN)	99.81%/0.18%
Varga et al. 2015	D	wykrycie palety określane są jako zależność sumy i przecięcia obszarów etykiety z detekcją, ale nieco inna niż w 2014	procent wykrytych palet (TP)	99.40%
Varga et al. 2016	D	jak w Varga et al. z 2015 roku	jak w Varga et al. z 2015 roku	100%

Tabela 2.2. Tabela z informacjami o ewaluacji oraz wynikach przeanalizowanych metod. L - lokalizacja, D - detekcja

2.1.5. Podsumowanie

Przeanalizowano 15 publikacji dotyczących detekcji palet na obrazie. Żadna z metod opartych na binaryzacji na podstawie barwy nie podejmuje istotnego problemu różnorodności kolorów

palet oraz natężeń oświetlenia. Oba te zjawiska występują w rzeczywistych warunkach w środowisku wózka widłowego i stanowią główne wyzwania. Metody zaproponowane przez autorów Cucciara R. oraz Varga R. oparte na cechach pozyskiwanych z obrazu w skali szarości adresują oba te problemy oraz wykazują dużą skuteczność detekcji w zróżnicowanych warunkach. Na szczególną uwagę zasługuje praca dr Varga R. oraz prof. Nedevschi S., która obejmuje 3 publikacje wychodzące w kolejnych latach 2014, 2015 i 2016. Autorzy przebadali rzetельnie kilka alternatywnych potoków przetwarzania, kilka algorytmów dla każdego z etapów przetwarzania oraz zbiór parametrów dla każdego z algorytmów. Ich metoda została wdrożona w projekcie PAN-Robots [35], a efekty są wykorzystywane obecnie w magazynach przemysłowych włoskiej firmy Elettric80.

W tabeli 2.3 zamieszczono podsumowanie przeanalizowanych metod pod kątem kluczowych elementów. Komentarz do tabeli oraz płynące z niej wnioski zamieszczono w sekcji 2.4 podsumowującej cały przegląd literaturowy w celu uchwycenia szerszego kontekstu.

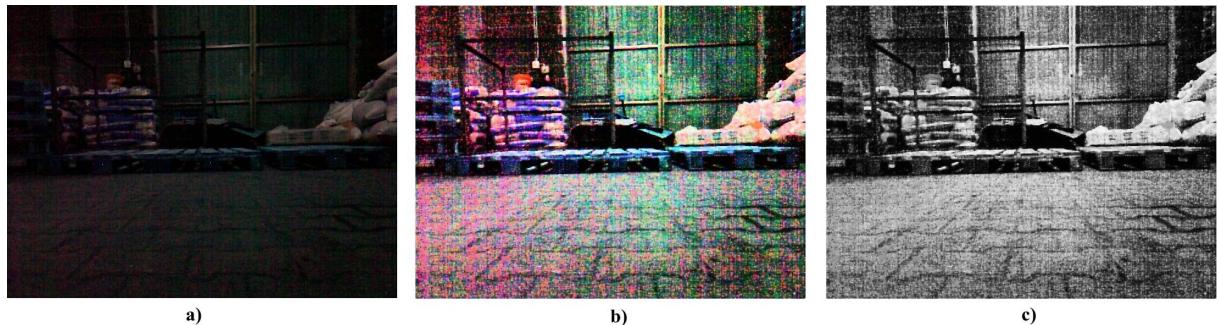
Metoda	Materiał, kolor i kształt palety	Wykrywany element	Zróżnicowane oświetlenie	Binaryzacja i rodzaj progu
Byun et al.	plastikowa, żółta, domknięta od dołu	front	nie	statyczne, ręcznie dobrane progi kanałów koloru
Chen et al	czerwono-biała, otwarta od dołu	front	nie	statyczne, ręcznie dobrane progi kanałów koloru
Syu et al.	drewniana, naturalny, domknięta od dołu	front	tak (ale ładunek zawsze zapewnia zacienienie dziur)	dynamiczny próg kontrastu w obszarze zainteresowania
Pages et al. i Cui et al.	drewniana, naturalny, domknięta od dołu	góra i oba boki	nie	statyczne, ręcznie dobrane progi kanałów koloru
Garibotto et al. z roku 1996 i 1997	domknięta od dołu	front	nie	nie
Kim et al.	plastikowa, czarna, domknięta od dołu	front	nie	nie
Cucciara et al.	drewniana, naturalny, otwarta od dołu	front	tak	nie
Varga et al. z 2014, 2015 i 2016 roku	drewniana, naturalny, otwarta od dołu	front	tak	nie

Tabela 2.3. Tabela z podsumowaniem przeanalizowanych metod detekcji.

2.2. ODSZUMIANIE OBRAZÓW

2.2.1. Wprowadzenie

Skuteczna metoda detekcji powinna działać możliwie niezależnie od warunków oświetleniowych. Warunki te są jednym z głównych problemów w przeanalizowanych w tym rozdziale publikacjach. Trudność w odróżnieniu palety od tła nie jest jedyną kwestią, która należy rozwiązać w warunkach ciemnych lub nadmiernie jasnych. Drugim problemem jest szum powstający w niedoświetlonych miejscach. Przykład takiego szumu na obrazie z wnętrza niedoświetlonego magazynu przemysłowego przedstawiono na Rys. 2.10. Na obrazie dokonano operacji wyrównania histogramu w celu zwiększenia kontrastu i uwydawnienia szumu. Dokonano również transformacji do obrazu w skali szarości, ponieważ jest to częsty zabieg na etapie przetwarzania wstępnego w poznanych metodach.



Rys. 2.10. Szum na obrazie wykonanym w niedoświetlonym magazynie przemysłowym; a) obraz oryginalny; b) obraz po wyrównaniu histogramu; c) obraz w skali szarości

Szum może utrudnić lub uniemożliwić skuteczne zastosowanie kolejnych etapów detekcji takich jak binaryzacja, segmentacja, detekcja krawędzi czy narożników. W celu skonstruowania metody detekcji opornej na słabe oświetlenie zapoznano się z dostępnymi metodami odszumiania.

2.2.2. Przegląd istniejących metod odszumiania

Dla problemu tak popularnego jak odszumianie obrazów tworzone są publikacje przeglądowe zawierające podsumowanie dorobku dziedziny w zakresie danego zagadnienia. Skorzystano z publikacji Fan et al. [17] z roku 2019, w której zaproponowano podział metod odszumiania na trzy główne kategorie takie jak metody klasyczne, transformacyjne oraz oparte o sieci neuronowe.

Metody klasyczne to najbardziej rozbudowana gałąź rozwiązań, którą można podzielić na filtry przestrzenne oraz metody bayesowskie. Popularnymi filtrami przestrzennymi są filtry liniowe takie jak filtr uśredniający [32] czy Wiener'a [5] oraz filtry nieliniowe takie jak filtr medianowy [32], ważony filtr medianowy [34] czy filtr bilateralny [42]. Głównym efektem filtrów przestrzennych jest rozmazanie, przez co nie nadają się one niestety do odszumiania obrazów, w których ważne jest zachowanie tekstuury. Z kolei metody bayesowskie można podzielić na regularyzację lokalną, nielokalną, reprezentację rzadką oraz minimalizację niskiego rzędu. Polegają one na optimalizacji pewnego wyrażenia, które reprezentuje prawdopodobieństwo otrzymania odszumionego obrazu. Kluczem do sukcesu jest dobranie wyrażenia tak, aby jego optymalizacja była jednoznaczna z oczekiwany rezultatem odszumienia. Najpopularniejszym reprezentantem metod regularyzacji

lokalnej jest regularyzacja TV [33], zakładająca, że obrazy są lokalnie wygładzone, a intensywność pikseli zmienia się stopniowo. Metoda jest efektywna i skutecznie zachowuje krawędzie, jednak czasem wygładza nadmiernie tekstury oraz zmienia kontrast obrazu. Najpopularniejszym reprezentantem metod regularyzacji nielokalnej jest natomiast algorytm średnich nielokalnych (NLM) [6]. Jego istotą jest szukanie podobnych regionów obrazu oraz uzależniania ich od siebie poprzez ważone uśrednianie ich wartości. Metoda ta jest bardziej złożona czasowo od regularyzacji lokalnych, za to wykazuje większą skuteczność na obrazach mocniej zaszumionych. Najpopularniejszym reprezentantem metod reprezentacji rzadkiej jest model nielokalnej scentralizowanej reprezentacji rzadkiej (NCSR) [16]. Stanowi on połączenie koncepcji reprezentacji rzadkiej wraz z regularyzacją nielokalną. Jest skuteczny zarówno na obszarach gładkich jak i zawierających tekstury, jednak jest jednocześnie wymagający obliczeniowo, co sprawia, że nie nadaje się do wielu zastosowań czasu rzeczywistego. Metody oparte o minimalizację rzędu dzielą się na metody polegające na faktoryzacji podobnych części obrazu do macierzy o mniejszym rzędzie oraz metody zmniejszania rzędu całego obrazu. Najpopularniejszą metodą jest ważona minimalizacja normy nuklearnej (WNNM) [21], która chociaż jest najbardziej skuteczną spośród wszystkich klasycznych metod, to jest również bardzo wymagająca obliczeniowo.

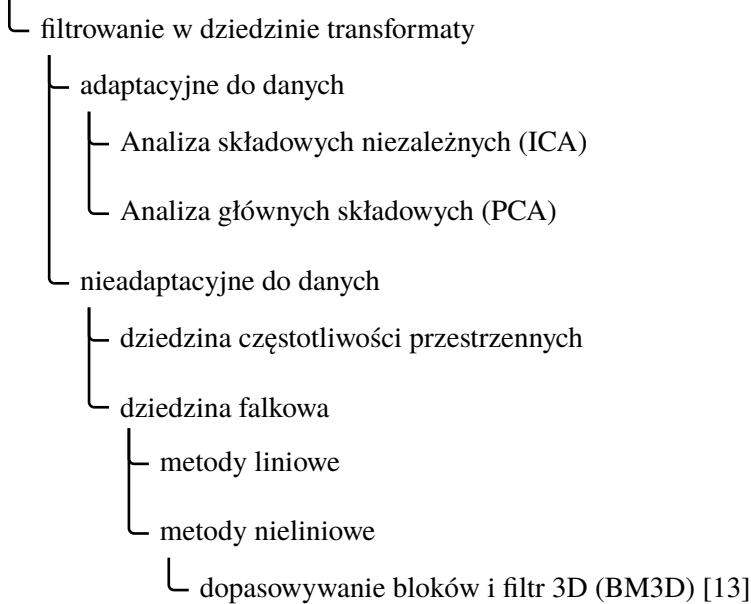
Drugą główną kategorię stanowią metody transformacyjne. Wśród nich dominuje podejście polegające na odróżnieniu szumu od obrazu w dziedzinie wybranej transformaty. Metody adaptacyjne do danych takie jak ICA oraz PCA wymagają próbek danych niezaszumionych oraz są wymagające obliczeniowo przez użycie okna przesuwnego, co sprawia, że są trudne do wykorzystania. Z kolei metody nieadaptacyjne do danych można podzielić ze względu na dwie dziedziny takie jak falkowa i częstotliwości przestrzennej. W dziedzinie częstotliwości przestrzennej stosuje się głównie filtry dolnoprzepustowe zakładające, że szum reprezentują wysokie częstotliwości, a obraz mniejsze. Z kolei dziedzina falkowa jest najczęściej badaną dziedziną w problemie odszumiania, ponieważ wykazuje dużą skuteczność w odseparowywaniu szumu niezależnie od jego częstotliwości. Wymaga ona jednak poprawnego dobrania hiperparametru nazywanego bazą falki. Z tego powodu uznawana jest za metodę nieadaptacyjną. Najlepszą metodą transformacyjną i najpopularniejszą metodą w całej dziedzinie jest metoda dopasowywania bloków i filtr 3D (BM3D) [13]. Stanowi ona silne rozszerzenie metody NLM i polega na pogrupowaniu podobnych fragmentów obrazu, a następnie poddaniu transformacji falkowej. W dziedzinie transformaty grupy poddawane są progowaniu lub filtrowi Wiener'a. Następnie stosuje się transformację odwrotną oraz odtwarza cały obraz na podstawie grup. Metoda ta nie radzi sobie z obrazami, w których szum narasta stopniowo, wprowadzając niechciane artefakty do obrazu.

Trzecią główną kategorię stanowią metody oparte o sieci neuronowe. Można podzielić je na sieci oparte o wielowarstwowy perceptron (MLP) będące głównie autokoderami oraz oraz sieci konwolucyjne (CNN) takie jak odszumiające konwolucyjne sieci neuronowe sprężone w przód (DnCNNs) czy szybka i elastyczna odszumiająca konwolucyjna sieć neuronowa (FFDNet) [52]. Przewagą tego typu metod jest krótszy czas działania niż iteracyjnie optymalizowanych metod bayesowskich, jednak kosztem bardzo długiego uczenia. Charakterystyczne dla sieci neuronowych jest wymaganie bardzo dużego zbioru danych w postaci par obrazów zaszumionego i niezaszumionego. Poważnym ograniczeniem jest fakt, że sieć wyuczona do zadania odszumiania pewnego konkretnego rodzaju szumu, najprawdopodobniej nie sprawdzi się do szumu z innego rozkładu.

Ze względu na dużą liczbę zagnieździeń w analizie rodzin metod odszumiania, opracowano oraz przedstawiono ich hierarchiczną strukturę na Rys 2.13 - 2.12. Ponadto ze względu na mnogość poznanych publikacji i algorytmów, wyłącznie wybrane rodziny zostały rozwinięte do przykładów konkretnych, kluczowych dla dziedziny metod. Szczegółowa analiza zawierająca

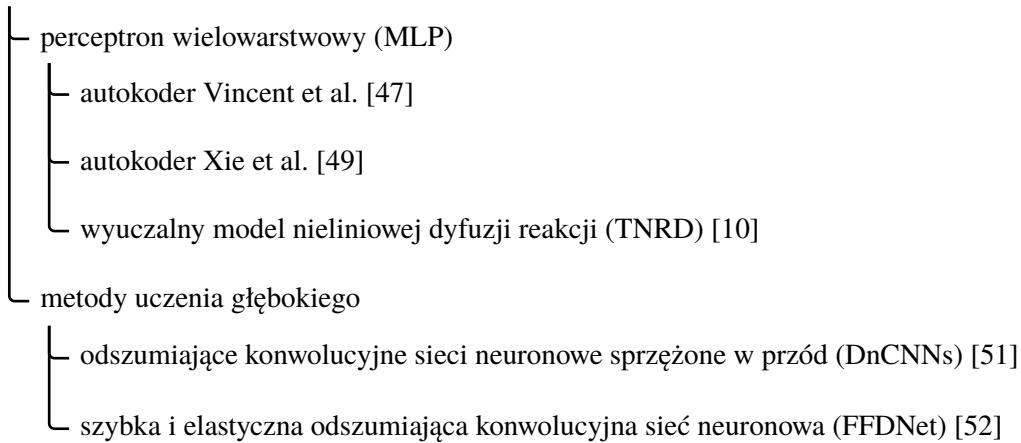
referencje do niszowych metod znajduje się w publikacji Fan et al. [17], na której opierają się przedstawiony podział.

Transformacyjne metody odszumiania obrazów



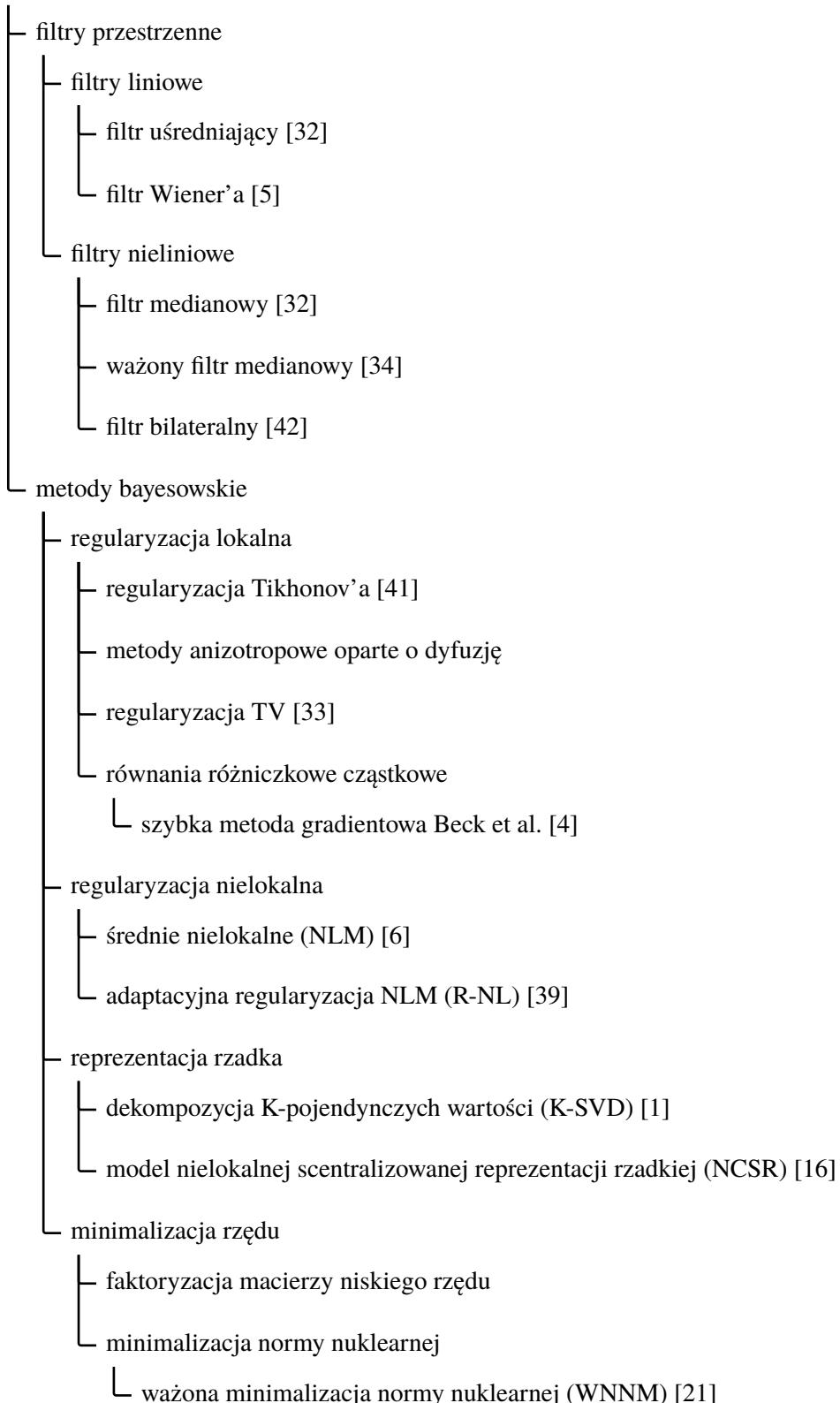
Rys. 2.11. Hierarchiczna struktura transformacyjnych metod odszumiania obrazów na podstawie publikacji przeglądowej Fan et al. [17]

Metody odszumiania w oparciu o sieci neuronowe



Rys. 2.12. Hierarchiczna struktura metod odszumiania obrazów w oparciu o sieci neuronowe na podstawie publikacji przeglądowej Fan et al. [17]

Klasyczne metody odszumiania obrazów



Rys. 2.13. Hierarchiczna struktura klasycznych metod odszumiania obrazów na podstawie publikacji przeglądowej Fan et al. [17]

2.2.3. Podsumowanie w odniesieniu do problemu detekcji palet

Po rozpoznaniu dostępnych metod oraz przeanalizowaniu ich zalet, wad, a także przeznaczenia należy zarysować charakterystykę problemu odszumiania obrazów w niedoświetlonych warunkach magazynów przemysłowych, a następnie wybrać metody, które są do niej najlepiej dostosowane. Pierwszą cechą algorytmów działających w czasie rzeczywistym jest szybkość działania. Jest to wymóg, którego najprawdopodobniej nie spełnią iteracyjnie optymalizowane metody bayesowskie oraz część metod transformacyjnych. Poznane szybkie metody to filtry przestrzenne oraz sieci neuronowe sprzążone w przód. Kolejnym wymogiem jest skuteczność dla szumu pochodzącego z nieznanego rozkładu wynikającego z ustawień parametrów kamery, którą pozyskiwany jest obraz. Szczególną wrażliwość na nieokreślony rozkład szumu wykazują metody oparte o sieci neuronowe. Pozostałe metody należałyby zbadać dla szumów z rozkładów innych niż najczęściej stosowany gaussowski. Kolejnym wymogiem może być skuteczne zachowywanie ostrych krawędzi oraz tekstury. Filtry przestrzenne wykazują szczegółowe własności rozmazujące, przez co na ogół nie spełniają tych warunków. Są to jednak wymogi silnie zależne od przyjętego algorytmu detekcji, ponieważ może okazać się, że przyjęta metoda ekstrakcji cech działa również skutecznie dla obrazów z ostrymi krawędziami i zachowaną tekstem jak i bez nich. Wszystkie przeanalizowane wymogi wskazują na dwie możliwe ścieżki postępowania. Pierwsza to wykorzystanie szybkich filtrów przestrzennych wraz z metodą detekcji niewrażliwą na rozmazanie krawędzi i tekstury wykrywanych obiektów. Druga to skorzystanie z metod transformacyjnych lub bayesowskich, które pozwalają zapobiec nadmierнемu rozmazaniu obiektów. Drugie rozwiązanie wymaga zaopatrzenia w sprzęt, który pozwoli na wykonywanie wymagających algorytmów w czasie rzeczywistym.

Przeanalizowano także publikacje przytoczone w sekcji 2.1 pod kątem występowania algorytmów odszumiania. W tabeli 2.4 zamieszczono podsumowanie. Komentarz do tabeli oraz płynące z niej wnioski zamieszczono w sekcji 2.4 podsumowującej cały przegląd literaturowy w celu uchwycenia szerszego kontekstu.

Metoda	odszumianie
Byun et al.	brak
Chen et al	brak
Syu et al.	brak
Pages et al. i Cui et al.	brak
Garibotto et al. z roku 1996 i 1997	brak
Kim et al.	brak
Cucciara et al.	brak
Varga et al. z 2014, 2015 i 2016 roku	filtr Gaussa

Tabela 2.4. Tabela z informacją o zastosowanym algorytmie odszumiania dla każdej z przeanalizowanych metod detekcji.

2.3. WYGŁADZANIE DETEKCJI W SEKWENCJI WIDEO

2.3.1. Wprowadzenie

Detekcja palet w środowisku wózka widłowego najczęściej odbywa się w trakcie najazdu autonomicznego wózka na miejsce, w którym spodziewana jest paleta. Miejsce to jest wskazane przez system GPS działający w magazynie. Dokładność takiej nawigacji to około pół metra.

Z tego powodu wymagana jest dużo dokładniejsza lokalizacja na podstawie obrazu z kamery przymocowanej do pojazdu. Kamera rejestruje sekwencję wideo na której spodziewamy się jednej lub kilku palet. Paleta jest obserwowana na kolejnych klatkach strumienia. Paleta znajduje się w tym samym miejscu względem poprzedniej klatki, gdy pojazd stoi w miejscu. Natomiast, gdy pojazd porusza się to wynik lokalizacji powinien przesuwać się na obrazie zgodnie z jego ruchem. Oczywiście jest, że każda z poznanych metod jest obarczona pewnym błędem. Na poszczególnych klatkach sekwencji wideo zbioru treningowego paleta może nie zostać wykryta, zostać wykryta w złym miejscu lub z przesunięciem. Ponadto paleta może być przez chwilę częściowo lub całkowicie przesłonięta innym obiektem. Obserwacja ta prowadzi do prostego wniosku, że wykorzystanie informacji o umiejscowieniu palety na poprzednich klatkach strumienia wideo może zostać wykorzystane do predykcji umiejscowienia na klatce bieżącej oraz kolejnych. Zadanie predykcji ruchu obiektu można rozwiązać kilkoma podstawowymi algorytmami, które opisano w tej sekcji. Są to estymacja pozycji na podstawie prędkości, filtr alfa, filtr alfa-beta, filtr alfa-beta-gamma, filtr Kalman'a, rozszerzony filtr Kalman'a oraz bezśadowy filtr Kalman'a.

2.3.2. Prosta estymacja

Najprostszą metodą predykcji jest **estymacja pozycji na podstawie prędkości**. Polega ona na obliczeniu prędkości v_t na podstawie różnicy pozycji obiektu na dwóch kolejnych klatkach. Pozycja obiektu na przyszłej klatce \hat{x}_t obliczana jest na podstawie prędkości oraz ostatniej pozycji (2.3). Metoda ta jest jednak bardzo wrażliwa na błędy pomiaru i nie dokonuje żadnej filtracji.

$$\hat{x}_t = x_{t-1} + v_{t-1}, \quad v_t = x_t - x_{t-1} \quad (2.3)$$

2.3.3. Filtry alfa

Nieco bardziej odporną na błędy pomiaru metodą jest **filtr alfa**. Zakłada on, że pozycja obiektu \hat{x}_t jest lokalnie stała. Estymacja pozycji jest równa poprzedniej estymacji (2.4) zaktualizowanej o ostatni błąd estymacji \hat{r}_t przemnożony przez współczynnik alfa (2.6). Błąd estymacji jest równy różnicy pomiędzy estymowaną pozycją, a pomiarem (detekcją) (2.5).

$$\hat{x}_t = \hat{x}_{t-1} \quad (2.4)$$

$$\hat{r}_t = x_t - \hat{x}_t \quad (2.5)$$

$$\hat{x}_t = \hat{x}_{t-1} + \alpha \hat{r}_t \quad (2.6)$$

Filtr alfa-beta [30] [36] jest rozszerzeniem filtru alfa o estymację prędkości obiektu \hat{v}_t , która jest z założenia lokalnie stała. Pozycja obiektu nie jest już wtedy lokalnie stała, lecz zmienia się zgodnie z prędkością. Estymowana pozycja jest równa sumie poprzedniej zaktualizowanej estymowanej pozycji oraz poprzedniej zaktualizowanej estymowanej prędkości (2.7). Pozycja i prędkość są aktualizowane w ten sam sposób co pozycja w filtrze alfa, z tą różnicą, że wprowadzamy nowy współczynnik beta dla prędkości (2.8).

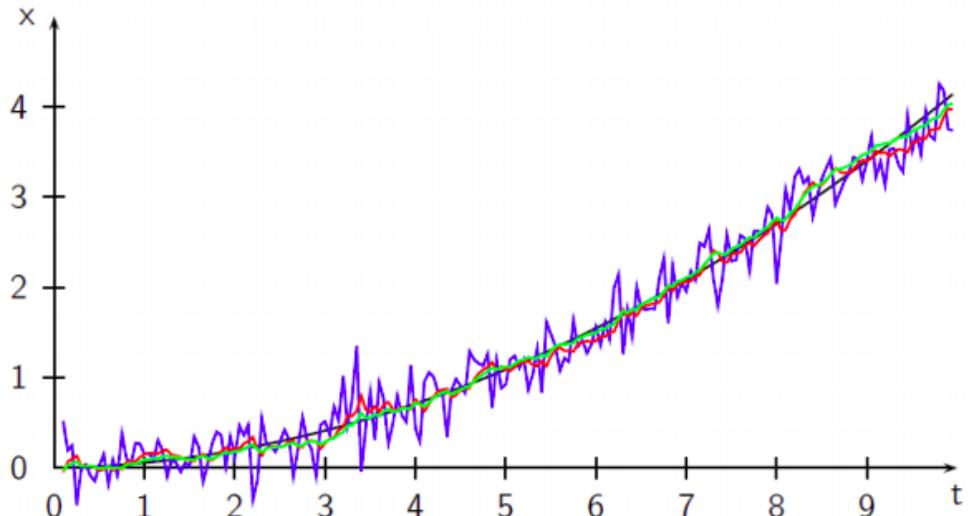
$$\hat{x}_t = \hat{x}_{t-1} + \hat{v}_{t-1}, \quad \hat{v}_t = \hat{v}_{t-1} \quad (2.7)$$

$$\hat{v}_t = \hat{v}_{t-1} + \beta \hat{r}_t \quad (2.8)$$

Kolejnym filtrem z tej rodziny jest **filtr alfa-beta-gamma**. Od filtra alfa-beta różni się on wprowadzeniem lokalnie stałego przyspieszenia oraz zmienną prędkością. Do wzoru na estymowaną pozycję dodajemy połowę poprzedniego zaktualizowanego przyspieszenia. Natomiast do wzoru na estymowaną prędkość dodajemy całą estymację przyspieszenia (2.9). Współczynnik gamma wprowadzony jest w celu aktualizowania przyspieszenia (2.10). Możliwe są wyższe rzędy filtrów z tej rodziny, jednak nie są one potrzebne do zamodelowania elektrycznego napędu autonomicznego pojazdu o stałym przyspieszeniu.

$$\hat{x}_t = \hat{x}_{t-1} + \hat{v}_{t-1} + \frac{1}{2}\hat{a}_{t-1}, \quad \hat{v}_t = \hat{v}_{t-1} + \hat{a}_{t-1}, \quad \hat{a}_t = \hat{a}_{t-1} \quad (2.9)$$

$$\hat{a}_t = \hat{a}_t + \gamma \hat{r}_t \quad (2.10)$$



Rys. 2.14. Porównanie filtrów alfa-beta oraz alfa-beta-gamma dla detekcji obiektu o ruchu jednostajnie przyspieszonym. Czarną linią oznaczono ruch jednostajnie przyspieszony, niebieską pomiar, czerwoną estymację filtru alfa-beta, a zieloną estymację filtru alfa-beta-gamma. [źródło: [36]]

2.3.4. Filtry Kalman'a

Kolejnym szeroko stosowanym filtrem jest **Filtr Kalman'a** [25][18]. Filtr ten przyjmuje bardziej sformalizowaną i ogólną postać. Jednak dla problemu wygładzania pozycji obiektu lokalizowanego na obrazie może przyjąć formę bliską filtrowi alfa i pokrewnym. W pełnym ujęciu proces poddawany filtracji przedstawiany jest w postaci wzorów stanu i wyjścia (2.11).

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) + v(t) \\ y(t) &= Cx(t) + w(t) \end{aligned} \quad (2.11)$$

gdzie:

- $x(t)$ = stan w chwili t
- $u(t)$ = wejście
- $y(t)$ = wyjście
- A = macierz tranzycji stanu
- B = macierz tranzycji wejścia
- C = macierz tranzycji wyjścia
- $v(t)$ = szum procesowy
- $w(t)$ = szum pomiarowy

Szum procesowy i pomiarowy są niezależne i pochodzą z rozkładu normalnego o znanych macierzach kowariancji V i W . Odwzorowują one niedoskonałość modelu oraz niedokładność pomiarów (2.12).

$$\begin{aligned} v(t) &\sim N(0, V) \\ w(t) &\sim N(0, W) \end{aligned} \quad (2.12)$$

Stan początkowy $x(0)$ również pochodzi z rozkładu normalnego. Wartości oczekiwana rozkładu stanu początkowego to x_0 , a macierz kowariancji to P_0 (2.13).

$$x(0) \sim N(x_0, P_0) \quad (2.13)$$

Algorytm filtru składa się z dwóch faz. Pierwsza związana jest z aktualizacją chwili w czasie i polega na obliczeniu kolejnej predykcji stanu $\hat{x}(t+1|t)$ wraz z macierzą kowariancji $P(t+1|t)$ (2.14).

$$\begin{aligned} \hat{x}(t+1|t) &= A\hat{x}(t|t) + Bu(t) \\ P(t+1|t) &= AP(t|t)A^T + V \end{aligned} \quad (2.14)$$

Druga faza polega na aktualizacji pomiarów w kolejnej chwili w czasie (2.15). W tym celu obliczana jest innowacja $\varepsilon(t)$, czyli różnica pomiędzy ostatnim pomiarem, a oszacowaniem stanu. Kluczową wartość stanowi również wzmacnienie Kalmana $K(t)$, które decyduje jaki wpływ na estymację ma nowy pomiar, a jaki poprzednia estymacja.

$$\begin{aligned} \varepsilon(t) &= y(t) - C\hat{x}(t|t-1) \\ S(t) &= CP(t|t-1)C^T + W \\ K(t) &= P(t|t-1)C^T s^{-1}(t) \\ \hat{x}(t|t) &= \hat{x}(t|t-1) + K(t)\varepsilon(t) \\ P(t|t) &= P(t|t-1) - K(t)S(t)K(t)^T \end{aligned} \quad (2.15)$$

Tak sformułowany filtr pozwala na zastosowanie w wielu procesach. W wersji podstawowej jest o jednak filtrem liniowym. Istnieją również rozszerzenia filtru takie jak **rozszerzony filtr Kalman'a** oraz **bezśladowy filtr Kalman'a** dla procesów nieliniowych. Rozszerzony filtr Kalman'a pozwala modelować nieliniowe procesy poprzez wykorzystanie funkcji w miejsce macierzy oraz linearyzację modelu. Bezśladowy filtr Kalman'a jest modyfikacją rozszerzonego filtru Kalman'a polegającą na odmiennym podejściu do wyznaczania parametrów zmiennej losowej stanu. Jednak jak wynika z przykładu w sekcji 2.3.5, filtr liniowy może być wystarczający do uzyskania skutecznego wygładzenia w problemie korelującym z detekcją palet w środowisku wózka widłowego.

2.3.5. Filtr Kalman'a w śledzeniu obiektów

W celu zapoznania z użyciem filtru Kalmana w problemie śledzenia obiektu na sekwencji wideo, przeanalizowano przypadek śledzenia ruchu osób opisany w pracy **Sheldon&Chang [50]**. Stan procesu zamodelowano w postaci czteroelementowego wektora zawierającego współrzędne obiektu na obrazie x, y oraz prędkości poruszania dx, dy (2.16).

$$x(t) = [x, y, dx, dy] \quad (2.16)$$

Macierz tranzycji stanu A jest skonstruowana tak, aby mnożenie przez wektor stanu realizowało proste dodawanie pozycji do prędkości dla obu współrzędnych (2.17). W tym miejscu zauważa się pełną analogię do wzoru na estymację pozycji w filtrze alfa-beta (2.7).

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.17)$$

Wejście $u(t)$ jest wartością skalarną i zawiera informację o przesunięciu na wzdłuż osi X, którą uzyskujemy z systemu kontroli robota. Macierz tranzycji wejścia B jest skonstruowana tak, aby mnożenie przez skalar wejścia $u(t)$ zwiększyło wymiarowość wejścia do wymiarowości wektora stanu, jednocześnie przekazując informację zawartą w wartości skalarnej (2.18).

$$B = [1 \ 0 \ 0 \ 0] \quad (2.18)$$

Kowariancja szumu procesowego V obliczna jest metodą Autocovariance Least-Squares na podstawie zbioru treningowego [31]. Kowariancja szumu pomiarowego W jest obliczana dynamicznie na podstawie dwóch ostatnich klatek oraz reguły w postaci równania liniowego. Sformułowanie reguły odbywa się na etapie uczenia i polega na znalezieniu cech obrazu, których zmiana wprowadza największy błąd pomiarowy, a następnie zastosowaniu regresji liniowej dla każdej pary następujących po sobie obrazów w zbiorze treningowym. Regresja liniowa pozwala znaleźć zależność pomiędzy cechami, a błędem, czyli szukaną regułę. Istnieje również prostsze podejście do wyznaczenia W . Można uprościć problem do obliczenia statystyk na podstawie zbioru treningowego bez użycia inżynierii cech i regresji liniowej.

Wyjście $y(t)$ to obie współrzędne wskazujące na pozycję obiektu na obrazie (2.19). W omawianej publikacji wyjście jest zmultiplikowane dla kilku badanych algorytmów, jednak na potrzeby prostego przykładu zostało to pominięte.

$$y(t) = [x, y] \quad (2.19)$$

Macierz tranzycji wyjścia C została sformułowana tak, aby po wymnożeniu przez stan procesu zwrócić dwie współrzędne wyjściowe (2.20).

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (2.20)$$

Tak sformułowany filtr Kalman'a został przebadany na różnych ujęciach poruszających się osób wraz z algorymem CamShift (Continuously Adaptive Mean Shift Algorithm) [2]. Pierwszy typ ujęć wyróżniał się tym, że tło i człowiek były trudne do rozróżnienia. Filtr pozwolił na zachowanie ciągłości lokalizacji w 23 procentach więcej przypadków niż algorytm bez filtra. Na ujęciach z drastycznie zmieniającymi się warunkami oświetleniowymi filtr nie wniosł żadnej poprawy. Natomiast na ujęciach, na których obiekt był rozmazany lub zacieniony, filtr pogorszył skuteczność śledzenia.

2.3.6. Podsumowanie

Przeanalizowano 7 metod wygładzania detekcji w sekwencji wideo. Każda z metod z wyjątkiem filtru Kalmana'a daje się opisać w kilku prostych równaniach. Filtr Kalman'a jako jedyna z rozpoznanych metod pozwala na uwzględnienie sygnałów wejściowych pochodzących z kontrolera robota. Wymaga on jednak sformułowania macierzowej reprezentacji procesu detekcji, co nie jest konieczne w przypadku pozostałych metod. Każda z metod daje szansę na podniesienie skuteczności detekcji oraz zwiększenie odporności na błędy powodowane chwilowym przesłonięciem obiektu na obrazie, bądź niepoprawną lub niedokładną detekcją.

Przeanalizowano także publikacje przytoczone w sekcji 2.1 pod kątem występowania filtrów do wygładzania detekcji. W tabeli 2.5 zamieszczono podsumowanie. Komentarz do tabeli oraz płynące z niej wnioski zamieszczono w sekcji 2.4 podsumowującej cały przegląd literaturowy w celu uchwycenia szerszego kontekstu.

Metoda	Filtr wygładzający po czasie
Byun et al.	brak
Chen et al	filtre Kalman'a
Syu et al.	filtre cząsteczkowy [27]
Pages et al. i Cui et al.	brak
Garibotto et al. z roku 1996 i 1997	brak
Kim et al.	brak
Cucciara et al.	brak
Varga et al. z 2014, 2015 i 2016 roku	filtre Kalman'a

Tabela 2.5. Tabela z informacją o zastosowanym filtrze wygładzaniu po czasie dla każdej z przeanalizowanych metod detekcji.

2.4. PODSUMOWANIE PRZEGŁĄDU

Po dokonaniu analizy istniejących metod w sekcji 2.1, algorytmów odszumiania w sekcji 2.2 oraz filtrów wygładzania po czasie w sekcji 2.3, zauważono pewne braki w dotychczasowych osiągnięciach badaczy.

Po pierwsze, istniejące rozwiązania dzielą się na dwie grupy dostrzegalne w tabeli 2.3. Metody w pierwszej grupie korzystają z barwy palety do binaryzacji, a metody w drugiej grupie korzystają wyłącznie z obrazów w skali szarości do pozyskania cech. Brakuje metody, która łączyłaby oba te rozwiązania, czyli wykorzystywała informację o barwie, lecz nie polegała wyłącznie na niej.

Po drugie, zauważa się braki w badaniu zagadnienia zróżnicowanych warunków oświetleniowych, a szczególnie niewidocznych, które powodują powstawanie charakterystycznego szumu opisanego w sekcji 2.2.1. Tabela 2.3 wskazuje, że zaledwie dwóch autorów testuje swoją metodę detekcji w środowisku prześwietlonym lub niewidocznym, z czego wyłącznie jeden stosuje filtr w celu odszumienia obrazów, jak pokazuje tabela 2.4.

Po trzecie, w większości prac zauważa się zupełny brak wykorzystania sekwencyjności analizowanych obrazów widoczny w tabeli 2.5. Natomiast nieliczne metody, które dokonują filtracji po czasie, wykorzystują złożone filtry w miejsce wcześniejszego zbadania prostszych rozwiązań np. z rodziny alfa, opisanych w sekcji 2.3.

Spośród trzech przytoczonych zagadnień pochyłono się nad pierwszym z nich, czyli stworzeniu zespołu metod, który wykorzystywałby informację o barwie palet oraz dokonywał ekstrakcji

cech niezależnych od niej, z obrazu w skali szarości. Autorska metoda inspirowana poznanymi rozwiązaniami została opisana w rozdziale 3.

3. OPIS BADANEJ METODY

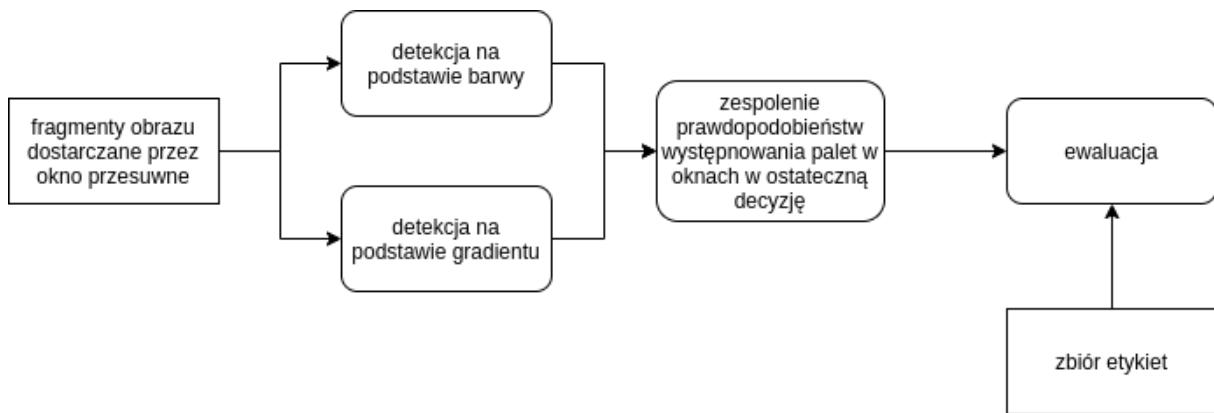
Celem rozdziału opisującego badaną metodę jest zapoznanie z opracowanym, autorskim sposobem, mającym za zadanie podniesienie skuteczności metod poznanych na etapie przeglądu literaturowego oraz wypełnienie najważniejszego zauważonych braków dotychczasowych badań, podsumowanych w sekcji 2.4. Dokładny opis metody jest niezbędny do zrozumienia i odtworzenia jej przez przyszłych czytelników oraz badaczy.

3.1. ARCHITEKTURA WYSOKIEGO POZIOMU

Opracowana w ramach niniejszej pracy metoda jest zasadniczo zespołem dwóch współpracujących metod. (Nie jest jasne czy parę metod można nazywać zespołem, jednak w niniejszej pracy wykorzystano to pojęcie w odniesieniu do połączenia dwóch metod, m.in. ze względu na powszechność pojęcia zespołu klasyfikatorów, które stosuje się do określenia złożenia nawet niewielkiej liczby klasyfikatorów składowych). Pierwsza z nich jest autorską metodą binaryzacji obrazu na podstawie koloru oraz wyliczania prawdopodobieństwa występowania palety we fragmencie obrazu. Z kolei druga metoda zapożyczona jest z pracy Varga et al. z roku 2015 [45] i wylicza prawdopodobieństwo na podstawie cech pozyskiwanych z obrazu w skali szarości, takich jak wielkość gradientu. Metoda Varga et al. 2015 została wybrana ze względu na wysoką skuteczność, przedstawioną w tabeli 2.2 oraz prosty potok przetwarzania, opisany w rozdziale 2.1.3. Obie metody zostały w pełni opisane w sekcjach 4.2.1 i 3.4. Obie metody przyjmują na wejście obraz w przestrzeni barw RGB, a na wyjściu zwracają prawdopodobieństwo występowania palety w danym fragmencie obrazu, dzięki czemu mogą pracować w połączeniu z techniką okna przesuwnego opisaną w sekcji 3.2. Każda z metod może działać osobno, dzięki czemu zbadana zostanie ich skuteczności względem zespołu. Natomiast głównym zadaniem zespołu, czyli połączenia obu tych metod, będzie podjęcie decyzji, na którym fragmencie obrazu jest paleta na podstawie prawdopodobieństw zwracanych przez obie metody składowe. Sposoby łączenia metod i podejmowania wspólnych decyzji zostały opisane szczegółowo w sekcji 3.5. Sposób ewaluacji każdej z metod składowych oraz zespołu został opisany w sekcji 3.6. Architekturę rozwiązania przedstawiono na rysunku 3.1

3.2. TECHNIKA OKNA PRZESUWNEGO

Technika okna przesuwnego jest jednym z najbardziej popularnych elementów metod detekcji obiektów na obrazie. Polega na podzieleniu obrazu na prostokątne obszary nazywane oknami o zmieniających się wymiarach oraz pozycji, które to okna poddawane są następnie ocenie czy zawierają szukany obiekt. Ocena odbywa się z wykorzystaniem konkretnej metody detekcji, której rezultatem jest wartość logiczna lub prawdopodobieństwo występowania obiektu, przyporządkowane każdemu wyznaczonemu obszarowi. Jeśli detekcja zwraca prawdopodobieństwa, to wymagane jest określenie sposobu w jaki na ich podstawie podejmujemy ostateczną decyzję, które obszary obrazu zawierają obiekt. Istnieje kilka sposobów określania zakresu pozycji oraz wymiarów prostokątnego obszaru. Częstą praktyką jest przyjęcie, że wymiary okna są stałe, a obraz na którym dokonywana jest detekcja, jest skalowany zgodnie z pewnym współczynnikiem. Należy



Rys. 3.1. Architektura wysokiego poziomu badanej metody detekcji wraz z ewaluacją.

także ustalić krok o jaki przesuwamy okno po obrazie horyzontalnie i wertykalnie. Ustawienie parametrów w sposób, który powoduje wyznaczenie dużej ilości okien, ma pozytywny wpływ na jakość detekcji, lecz negatywny na czas detekcji, stąd nierzadko wymagane jest przyjęcie pewnego kompromisu pomiędzy tymi dwoma aspektami. Wizualizacje okna przesuwnego przedstawiono na rysunku 3.2.



Rys. 3.2. Wizualizacja pojedynczego okna wyznaczonego techniką okna przesuwnego; a) kolorem zielonym zaznaczono pojedyncze okno; b) obszar obrazu wycięty przez zaznaczone okno

W ramach badań opisanych w rozdziale 4 przyjęto następujące wartości parametrów:

$$\begin{aligned}
 \text{krok przesunięcia w obu kierunkach} &= 4 \text{ px} \\
 \text{współczynnik skalowania obrazu} &= 1.15 \\
 \text{wymiary okna} &= 22 \times 120 \text{ px}
 \end{aligned}$$

Optymalizacja skuteczności detekcji ze względu na wartości parametrów okna przesuwnego nie stanowi przedmiotu badań niniejszej pracy. Ważne jest wyłączne ustalenie niezmiennych parametrów, na których porównane zostaną metody poznane na etapie przeglądu z samodzielnie zaproponowanymi.

Częstym zjawiskiem występującym podczas predykcji techniką okna przesuwnego jest rozpoznanie przez klasyfikator obiektu w kilu częściowo pokrywających się oknach. Aby temu zapobiec

zastosowano pewne ograniczenie. Po wybraniu okna z największym prawdopodobieństwem występowania palety, odrzucane są wszystkie okna pokrywające się z nim w stopniu większym niż 0.33 zgodnie z miarą indeksu Jaccarda (opisaną szerzej w sekcji 3.6 dotyczącej ewaluacji). Wartość 0.33 występuje, gdy część wspólna dwóch okien o tych samych rozmiarach stanowi połowę jednego z nich. Po odrzuceniu tych okien można wybrać kolejne okno z największym prawdopodobieństwem i ponówić odrzucenie pokrywających je okien.

3.3. DETEKCJA W OPARCIU O BARWE

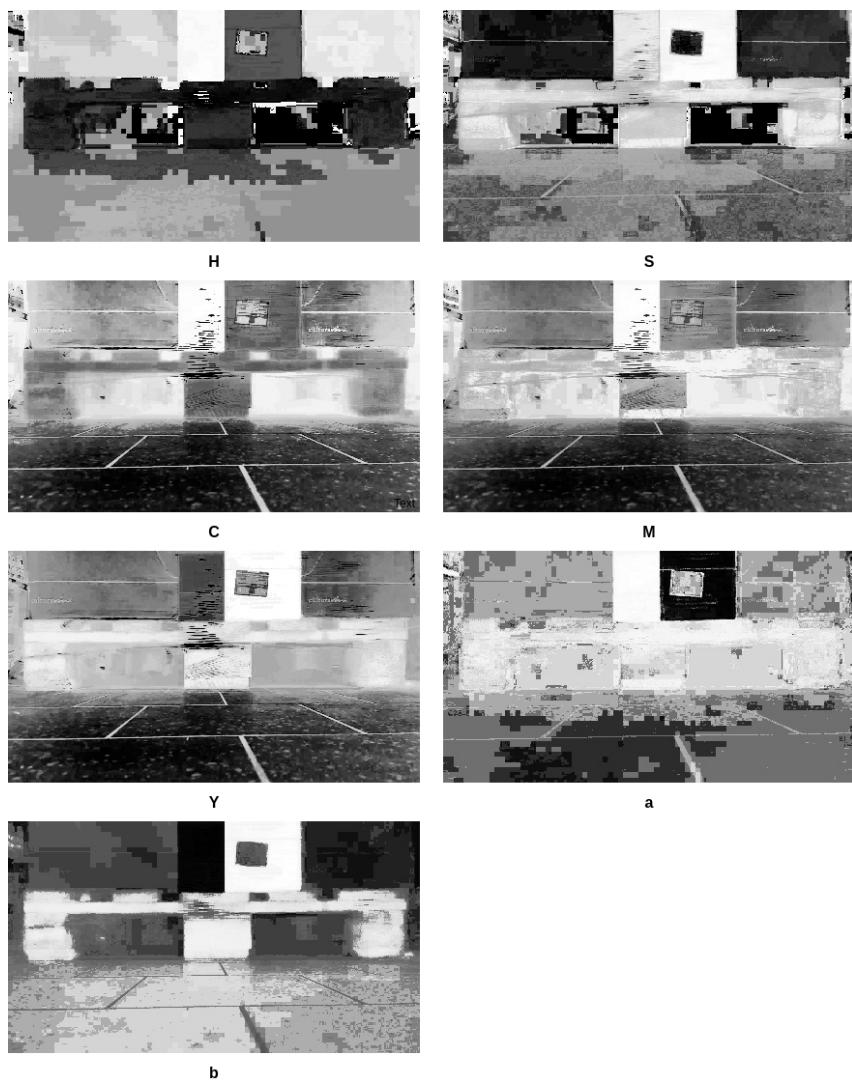
Pierwszym typem detekcji, który składa się na zespół metod badany w ramach pracy, jest binaryzacja w oparciu o Naiwny klasyfikator bayesowski z rozkładem Gaussa pracujący na cechach dotyczących barwy palet. Inspiracją do skonstruowania klasyfikatora są podejścia omówione w sekcji 2.1.2 przeglądu literaturowego, zawierającej metody dokonujące binaryzacji głównie na podstawie barwy. Są to metody, w których badania przeprowadzane są na paletach o ustalonym, niezmiennym, czasem jaskrawym kolorze. W pracy Byun et al. [7] - kolor żółty, Chen et al. [9] - kolor czerwony oraz Pages et al [29] i Cui et al. [22] - kolor naturalnego drewna. Dominującym założeniem wspólnym dla przytoczonych metod jest łatwa rozróżnialność koloru palety od tła. Wynikiem ich oddzielenia jest obraz binarny, na którym paleta namalowana jest białymi pikselami, a tło czarnymi. Decyzja, czy dany piksel należy do palety czy tła, odbywa się poprzez przyjęcie pewnych ręcznie ustalonych i niezmiennych progów wartości dla kanałów koloru. Jest to podejście mało elastyczne i wysoce podatne na zmiany takie jak natężenie oświetlenia, fabryczne różnice w odcieniach plastikowych palet czy gatunek drewna i sposób wykończenia w przypadku palet drewnianych.

W celu rozwiązania tych problemów zaproponowano Naiwny klasyfikator bayesowski modelujący rozkład Gaussa, który dla każdego kanału koloru uczy się średniej oraz wariancji jego wartości. Pozwala to na wyeliminowanie czynnika ludzkiego, a jednocześnie oparcie binaryzacji na wnioskowaniu na podstawie zbioru treningowego. Ponadto w celu uniezależnienia się od warunków oświetleniowych, stosowana jest trywialna inżynieria cech polegająca na transformacji profilu RGB na przestrzenie, w których barwa i kontrast stanowią osobne kanały. Są to np. HSV, CMYK czy Lab, w których kanały HS, CMY oraz ab odpowiadają za barwę, a V, K oraz L za kontrast. Kanały kontrastu są porzucane. Każdy kanał barwy poddawany jest operacji wyrównania histogramu w celu uwydatnienia potencjalnych różnic wartości między paletą a tłem. Wizualizację wykorzystywanych kanałów przedstawiono na Rys. 3.3. W proponowanym klasyfikatorze kanały barwy stanowią więc cechy piksela, który poddawany jest klasyfikacji na wspomniane klasy "paleta" i "tło".

Zbiór treningowy i testowy potrzebny do nauczenia oraz testowania jakości klasyfikacji pikseli pozyskiwany jest z obrazów w zbiorze danych poprzez wycięcie oetykietowanych palet, odrzucenie pikseli należących do dziur na widły oraz utworzenie listy pikseli należących do wszystkich palet ze wszystkich ujęć. Pozostałe piksele nienależące do palet uznawane są za piksele tła i również zostają skonkatenowane w jedną listę, która jednak ze względu na duży rozmiar może być próbkowana. Szczegóły dotyczące zbioru danych wykorzystanego do badań opisano w sekcji 4.1.

Zastosowanie Naiwnego klasyfikatora bayesowskiego nie rozwiązuje jednak kolejnego problemu, z którym borykają się wszystkie rozpoznane w ramach przeglądu metody, czyli detekcji nie tylko jednego, ale kilku kolorów palet jednocześnie. Problem ten rozwiązano poprzez skonstruowanie kilku klasyfikatorów dla każdego koloru występującego w zbiorze danych z osobna.

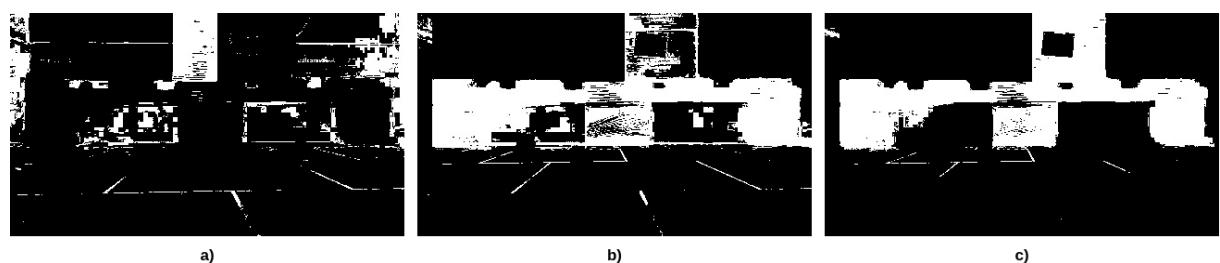
Trywialnie prosta budowa klasyfikatora oraz wysoka szybkość działania sprawiają, że predykcja, za pomocą kilku klasyfikatorów jednocześnie, nie jest obciążająca wydajnościowo. Takie



Rys. 3.3. Wizualizacja kanałów koloru wykorzystywanych do klasyfikacji na przykładowym obrazie z paletą. Przykłady uwzględniają operację wyrównania histogramu.

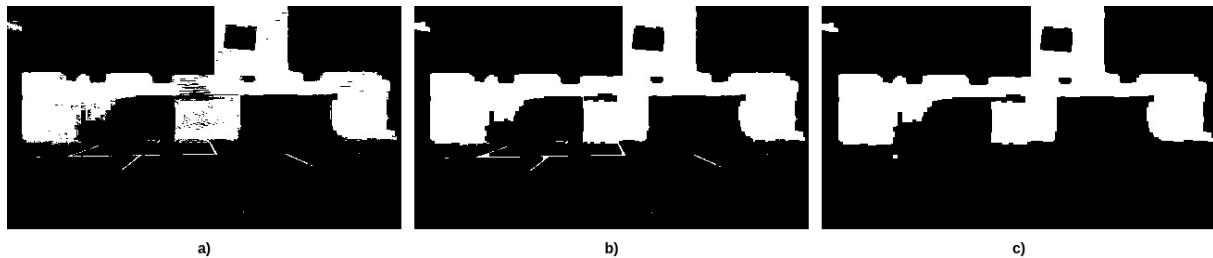
rozwiązanie wymaga jednak określenia sposobu łączenia wyniku dla kilku klasyfikatorów koloru. Aspekt ten został opisany pod koniec sekcji.

Wytrenowany klasyfikator służy do klasyfikacji każdego piksela zdjęcia z osobna, której wynikiem jest obraz binarny przedstawiony na Rys. 3.4.



Rys. 3.4. Rezultat binaryzacji obrazu zawierającego paletę w kolorze jasnego drewna, z wykorzystaniem klasyfikatorów wytrenowanych na trzech kolorach; a) niebieskim; b) ciemnego drewna; c) jasnego drewna

W celu poprawienia jakości binaryzacji stosowane są proste operacje morfologiczne będące kolejno domknięciem oraz otwarciem. Operacja otwarcia polega na zastosowaniu kolejno erozji i dylatacji, których wynikiem jest usunięcie szumu w postaci samotnych lub odstających, niewielkich grup pikseli, które zostały błędnie zaklasyfikowane jako piksele palety. Natomiast operacja domknięcia jest odwrotnością operacji otwarcia w sensie kolejności, a jej rezultatem jest wypełnienie dziur w obrębie palety, które błędnie zostały zaklasyfikowane jako tło, a otoczone są znaczącą liczbą pikseli palety. Rezultat operacji morfologicznych przedstawiono na Rys. 3.5.



Rys. 3.5. Rezultat operacji morfologicznych; a) wejściowy obraz binarny; b) wynik operacji domknięcia; c) wynik operacji domknięcia i otwarcia

Ostatnim etapem potoku przetwarzania zaproponowanej metody jest integracja z techniką okna przesuwnego. Każdy fragment obrazu binarnego wyciętego przez okno oceniany jest pod kątem prawdopodobieństwa występowania w nim palety. W celu określenia w jakim stopniu zawartość okna odpowiada wzorcowej masce palety, obliczana jest wartość funkcji dopasowania przyjmującej następującą postać:

$$(P_1 + D_0)/M \quad (3.1)$$

gdzie:

P_1 = liczba pikseli palety (białych) w obszarze palety

D_0 = liczba pikseli tła (czarnych) w obszarze dziur

M = łączna liczba pikseli okna bez marginesów

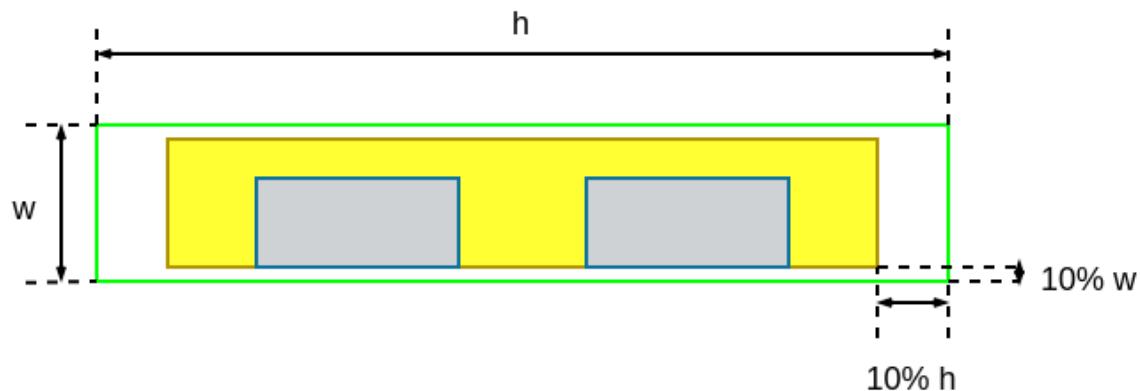
Pojęcia "obszar palety" i "obszar dziur" oraz marginesy preczyje rysunek 3.6.

Rozwiążanie wykorzystujące kilka klasyfikatorów, dla każdego koloru z osobna, wymaga określenia sposobu łączenia wartości prawdopodobieństwa występowania palety w danym oknie zawierającym fragment obrazów binarnych będących wynikiem klasyfikatorów kilku kolorów. W niniejszej pracy jako metodę agregacji przyjęto maksimum. Wybór został dokonany dla uproszczenia intuicyjnie.

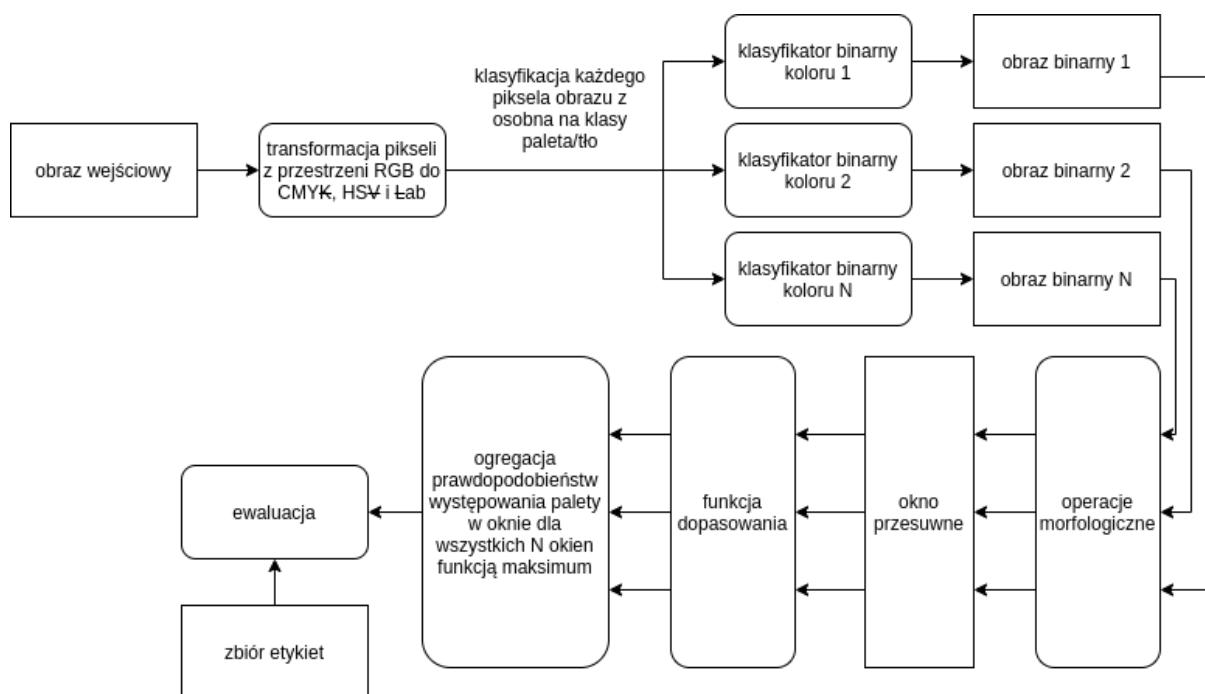
Opisany potok przetwarzania całej metody przedstawiono na rysunku 3.7. Zarówno klasyfikatory dla kolorów palet jak i całą metodę detekcji zbadano na autorskim zbiorze danych opisanym w sekcji 4.1. Wyniki przedstawiono w sekcji 4.2 w rozdziale dotyczącym badań i eksperymentów,

3.4. DETEKCJA W OPARCIU O GRADIENT

Drugie podejście składające się na zespół metod badany w ramach pracy inspirowane jest publikacjami Varga et al. z roku 2014 [44], 2015 [45] oraz 2016 [46], opisanymi szerzej w sekcji



Rys. 3.6. Okno przesuwne z zaznaczonymi obszarami wymaganymi do obliczenia funkcji dopasowania 3.1 oraz marginesami; w - szerokość okna, h - wysokość okna, zielony prostokąt - okno przesuwne; żółte wypełnienie - obszar palety; szare wypełnienie - obszar dziur



Rys. 3.7. Potok przetwarzania metody detekcji w oparciu o barwę

2.1.2 przeglądu literaturowego, zawierającej metody oparte na ekstrakcji cech z obrazów w skali szarości.

Potok przetwarzania rozpoczynany jest od konwersji obrazu z przestrzeni RGB do skali szarości. Zabieg ten pozwala uniezależnić się od barwy palety, a tym samym sprawia, że klasyfikator pracujący na cechach z obrazu w skali szarości będzie bardziej uniwersalny. Kolejnymi krokami przetwarzania wstępnego są zastosowanie filtra Gaussa, który wygładza krawędzie na obrazie oraz zamazuje detale, a następnie operacja wyrównania histogramu, która łagodzi wpływ niekorzystnych i niejednorodnych warunków oświetleniowych oraz zwiększa wartość gradientu. Tak przetworzony obraz w skali szarości stanowi pierwszy z ośmiu kanałów cech.

Kolejny kanał wyznaczany jest poprzez zastosowanie filtra Sobel'a dla osi X oraz Y, na podstawie którego możemy obliczyć wielkość gradientu na obrazie. Ostatnie sześć kanałów również zawiera wielkość gradientu jednak zorientowanego, z sześciu różnych zakresów kątów. Są to 0-60, 60-120, 120-180, 180-240, 240-300 oraz 300-360 liczone w stopniach. Tak przygotowane osiem kanałów jest spłaszczone do jednego wektora, stanowiącego cechy obrazu podawane na wejście klasyfikatora. Rezultaty wszystkich etapów przetwarzania prowadzące do ekstrakcji cech przedstawiono na Rys. 3.8.

Zbiór danych uczących oraz testowych tworzony jest poprzez wycięcie ze wszystkich obrazów prostokątów o zadanych rozmiarach odpowiadających oknu przesuwnemu. W wyciętym prostokącie powinna mieścić się wykrywana paleta oraz 10 procent marginesu wokół niej. Należy w ten sposób wyciąć ze zbioru danych wszystkie palety oraz co najmniej drugie tyle obszarów zawierających tło.

Klasyfikator zbadany w pracy Varga et al. z roku 2014 [44] to zespół AdaBoost z drzewami decyzyjnymi w liczbie 2048. Zauważono jednak, że klasyfikator nazywany Lasem Losowym jest rozwiązaniem bardziej efektywnym pamięciowo oraz obliczeniowo, z jednoczesnym zachowaniem wysokiej skuteczności. Do badań w niniejszej pracy magisterskiej wykorzystano więc Las Losowy o rozmiarze 128 drzew.

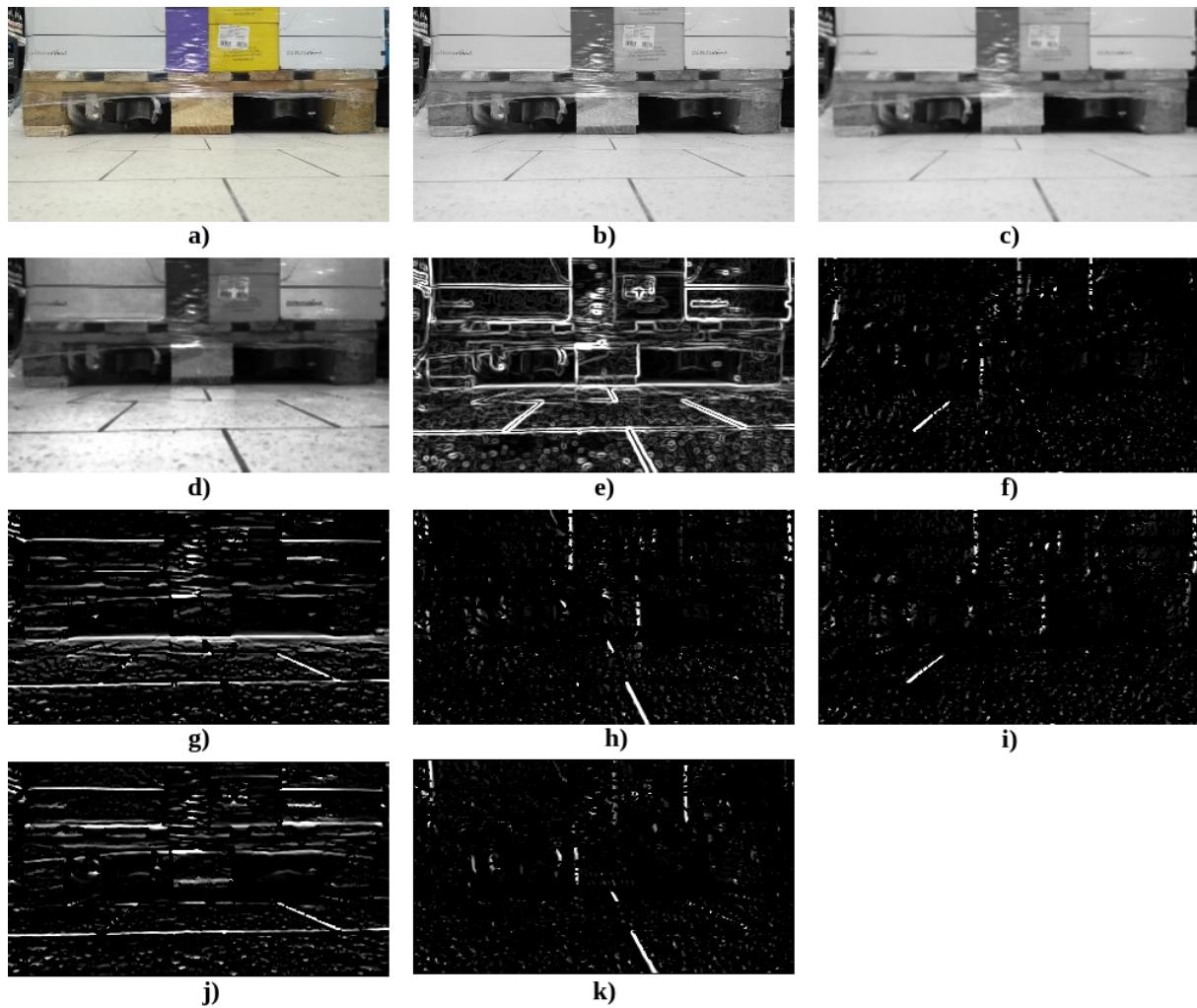
Zarówno klasyfikator Lasu Losowego jak i całą metodę detekcji połączoną z techniką okna przesuwnego zbadano na autorskim zbiorze danych opisanym w sekcji 4.1. Wyniki przedstawiono w sekcji 4.2 w rozdziale dotyczącym badań i eksperymentów.

3.5. ZESPÓŁ METOD DETEKCJI

Gdy dysponujemy dwiema metodami detekcji opisanymi w sekcjach 4.2.1 oraz 3.4, ostatnim wyzwaniem staje się połączenie obu metod w zespół, który pozwoli na podniesienie skuteczności detekcji. Autorska innowacja polega na wykorzystaniu jednocześnie informacji o kolorze palety jak i cech, które są niezwiązane z kolorem, takich jak kontrast i wielkość gradientu oraz jego orientacja. Łączenie odbywa się na etapie, w którym dysponujemy dwiema wartościami prawdopodobieństwa występowania palety w każdym oknie obrazu. Dwie wartości dostarczane są przez dwie metody, po jednej każda. Należy wybrać sposób agregowania obu wartości do jednej, na podstawie której następnie podejmiemy ostateczną decyzję, które okna zawierają palety. Zaproponowano metody agregacji: suma, maksimum, minimum oraz suma ważona. Każdą z metod przebadano także w wariancie ze znalezionymi wartościami.

Suma jest najbardziej intuicyjnym sposobem agregacji. Możemy ją interpretować jako połeganie takiego samego zaufania w obu wartościach prawdopodobieństwa zwracanych przez metody składowe. W tym sposobie agregacji zdecydowaną przewagę uzyskują okna, dla których obie wartości były wysokie. Dokładnie ten sam rezultat można uzyskać obliczając średnią, dlatego została ona pominięta.

Możemy również przyjąć założenie, że zaufamy wyłącznie jednemu klasyfikatorowi na raz, a więc zagragujemy obie wartości poprzez wybranie ich maksimum. Może się niestety zdarzyć, że wartości prawdopodobieństw będą pochodziły z nieco innego rozkładu. Taka sytuacja wystąpi, gdy na przykład funkcja oceny w metodzie wykorzystującej kolor będzie zwracała wartości prawdopodobieństwa z rozkładu o średniej 0.8, a Las Losowy zastosowany w drugiej metodzie będzie zwracał wartości z rozkładu o średniej 0.4. W takiej sytuacji metoda, która ma wyższą średnią wartości stanie się dominującą, gdy agregujemy poprzez maksimum. Aby zapobiec takiej sytuacji zaproponowano normalizację wartości. Typem zastosowanej normalizacji jest standaryzacja, która polega na odjęciu średniej od wartości, a następnie podzieleniu przez



Rys. 3.8. Etapy przetwarzania obrazów w metodzie gradientowej; a) obraz wejściowy; b) obraz w skali szarości; c) obraz po nałożeniu filtra Gaussa; d) obraz po wyrównaniu histogramu; e) wielkość gradientu; od f) do k) zorientowana wielkość gradientu w sześciu zakresach kątów podanych w stopniach, gdzie zakresy to kolejno 0-60, 60-120, 120-180, 180-240, 240-300 oraz 300-360

odchylenie standardowe, jak na wykresie 3.2.

$$z = \frac{x-\mu}{\sigma} \quad (3.2)$$

gdzie:

z = zmienna standaryzowana

x = zmienna niestandardaryzowana

μ = średnia z populacji

σ = odchylenie standardowe populacji

populacja = wartości prawdopodobieństwa palety dla wszystkich okien i obrazów

Taki zabieg sprawi, że prawdopodobieństwa zwracane przez obie metody staną się wartościami o średniej 0 i odchyleniu standardowym równym 1, co zapobiegnie dominacji jednej z nich. W przypadku maksimum, zabieg ten można interpretować jako wyrównanie szans przed swoistą bitwą obu metod o to, która zwróci największe prawdopodobieństwo.

Z kolei metoda minimum promuje okna, dla których wartości obu prawdopodobieństw są wysokie. Warto jednak zauważyć, że suma również posiada tę własność, a nawet więcej, ponieważ nie ignoruje drugiej wartości. Z tego powodu skuteczność agregacji największego minimum spodziewana jest być niższa niż prostej sumy. Dla tego sposobu agregacji również warto zbadać wpływ normalizacji.

Najbardziej adaptacyjnym do danych sposobem agregacji jest suma ważona, która polega na znalezieniu wag dla obu prawdopodobieństw oraz wymnożenie ich z wartościami przed sumowaniem. Takie podejście wymaga zbudowania zbioru par wag z pewnego zakresu, które zostaną poddane ewaluacji. Jest to podejście najbardziej czasochłonne i może wymagać wydzielenia reprezentatywnego, a jednocześnie nie zbyt wielkiego zbioru walidacyjnego. W celu uproszczenia zadania i pominięcia tego procesu zaproponowano wagę o wartościach równych skuteczności danej metody składowej wyrażonej indeksem Jaccarda. Za takim podejściem stoi intuicja polegająca na tym, że ufność jaką pokładamy w danej metodzie jest równa jej skuteczności.

3.6. EWALUACJA

Ewaluacja obu metod detekcji palet opisanych w sekcjach 3.4 oraz 4.2.1, a także zespołu opisanego w sekcji 3.5, wymaga trzech elementów.

Pierwszy element to ręcznie wyznaczone etykiety, będące prostokątami okalającymi palety na obrazach w zbiorze danych. Proces etykietowania opisany został szerzej w podsekcji 4.1.2 w sekcji eksperymentów dotyczącej zbioru danych.

Drugim elementem są prawdopodobieństwa występowania palety przypisane do każdego prostokąta wyznaczonego techniką okna przesuwnego. Obie metody umożliwiają pozyskanie takiej wartości dla obszaru o zadanych rozmiarach wyciętego ze zdjęcia. Metoda detekcji w oparciu o kolor oblicza prawdopodobieństwo na podstawie funkcji dopasowania 3.1. Natomiast metoda gradientowa udostępnia prawdopodobieństwo zwarcane przez klasyfikator lasu losowego.

Trzecim elementem jest proces wyznaczenia obszarów na obrazie z największym prawdopodobieństwem. Liczba palet na obrazie znana jest z góry. Wynika to z faktu, że system wizyjny mocowany w autonomicznych wózkach widłowych służy do precyzyjnego określenia lokalizacji palety, której współrzędne są znane dzięki współpracującemu z nim systemowi nawigacji w magazynie. System nawigacji obarczony jest błędem uniemożliwiającym bezpieczne wjechanie wideł pod palety, jednak wystarcza do wskazania robotowi, w którym miejscu magazynu znajdują się palety. Założenie to obecne jest w pracy Varga et al. z roku 2014 [44], w sekcji "C. Detection

postprocessing" w rozdziale "Proposed approach", która znalazła zastosowanie we włoskich zautomatyzowanych magazynach.

Gdy opracujemy te trzy elementy możemy dokonać ewaluacji. Na podstawie analizy dotychczas wykorzystanych metod ewaluacji w ramach przeglądu literaturowego w sekcji 2.1.4 zauważono, że żadna z poznanych metod nie pozwala na badanie jakości detekcji niezależne od rozdzielczości obrazu i ręcznie przyjętych progów dopuszczalnego błędu. W celu wykonania badań na obrazach o różnej rozdzielczości oraz uniknięcia wprowadzania nowych, bezpodstawnych wartości progów, zaproponowano wykorzystanie popularnej metody ewaluacji skupiającą się na precyzyji detekcji, nazywanej indeksem Jaccarda czy z języka angielskiego Intersection over Union (IoU). W przypadku detekcji polega ona na policzeniu współczynnika pokrywania się obszaru ręcznie wyznaczonego oraz wykrytego dla danej palety. Procent pokrywania wyliczany jest poprzez podzielenie pola obszaru części wspólnej dwóch prostokątów przez ich sumę 3.3:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3.3)$$

gdzie:

J = indeks Jaccarda

A = pole wypredykowanego prostokąta

B = pole prostokątnej etykiety

Przykładowe wartości współczynnika pokrywania wyznaczanego indeksem Jaccarda dla wybranych predykcji przedstawiono na rysunku 3.9. Ostateczne pokrywanie będące wynikiem całej ewaluacji jest uśrednionym pokrywaniem po wszystkich paletach w zbiorze danych testowych.



Rys. 3.9. Przykładowe wartości indeksu Jaccarda dla wybranych predykcji. Kolor niebieski to etykieta, kolor zielony to predykcja.

4. BADANIA I EKSPERYMENTY

4.1. ZBIÓR DANYCH

4.1.1. Pozyskanie i cechy

W celu przeprowadzenia badań zbudowano zbiór danych zawierający nagrania z frontalnych najazdów na palety, które symulują widok z kamery w trakcie poruszania się autonomicznego wózka widłowego. Nagrania wykonano w magazynie przemysłowym oraz hipermarkecie. Pozy-skano łącznie 29 ujęć, których suma pojedynczych klatek wynosi 2264. Część ujęć zawierają po kilka palet, stąd łączna liczba pojedynczych palet występujących w całym zbiorze przewyższa liczbę klatek i równa jest 2591. Oko kamery na każdym ujęciu znajduje się na wysokości kilku centymetrów ponad podłożem, czyli mniej więcej w połowie wysokości frontu palety. Obserwowany typ palety to popularna paleta euro EPAL o wymiarach 1200x800x144 mm. W zbiorze występują trzy kolory palet. Są to naturalne drewno, bejcowana (ciemna) oraz niebieska. Ponadto część kamerowanych palet zawiera ładunek i otoczona jest przezroczystą folią typu "stretch" służącą do zabezpieczania towarów. Nagrania odbywały się w zróżnicowanych warunkach oświetleniowych. Filmy rejestrowano w 4 rozdzielczościach takich jak HD (1280,720), VGA (640,480), 13MP (4160,3120) oraz 8MP (3264,2448). Przykładowe zdjęcia ze zbioru danych zostały przedstawione na Rys. 4.1.

4.1.2. Etykietowanie

W celu zbudowania dowolnego modelu dokonującego detekcji, należy nie tylko przygotować zbiór danych, ale również nadać etykiety każdemu zdjęciu, które reprezentować będą wzorcową detekcję, to znaczy wykonaną przez człowieka. W przypadku problemu detekcji obiektu na obrazie, na etykietę składają się współrzędne oraz wymiary obszarów wskazujących na szukany obiekt. W problemie detekcji palet będą to prostokąty okalające fronty palet. Proces etykietowania przeprowadzono z wykorzystaniem narzędzia napisanego przez autora w języku Python, które pozwala na przeskalowanie każdego zdjęcia do rozmiaru ułatwiającego ręczne zaznaczenie prostokątnego obszaru oraz zapis współrzędnych wyznaczonych obszarów, pogrupowanych na zdjęcia (klatki) i sceny. Zadanie etykietowania jest czasochłonne i wymaga precyzji, która może jednak zaowocować większą skutecznością modeli. Przykłady oetykietowanych zdjęć zostały przedstawione na Rys. 4.2.

4.2. BADANIA METOD SKŁADOWYCH

Eksperymenty rozpoczęto od przebadania metod detekcji w oparciu o barwę i gradient, które stanowią części składowe zaproponowanego zespołu metod. Przebadano zarówno skuteczność całych metod indeksem Jaccarda, w celu późniejszego porównania ich do zespołu, jak i klasyfikatorów pracujących w ramach obu metod, których skuteczność wykazano najbardziej popularnymi miarami: accuracy, precision, recall i F-score. Wszystkie badania przeprowadzono w formie



Rys. 4.1. Przykładowe klatki z ujęć ze zbioru danych. Dwa ostatnie zdjęcia są wynikiem operacji wyrównania histogramu, która ułatwia ludzkiemu oku dostrzeżenie palet zarejestrowanych w ciemnych warunkach oświetleniowych.



Rys. 4.2. Przykładowe klatki z ujęć ze zbioru danych z Rys. 4.1 z naniesionymi etykietami.

walidacji krzyżowej dzieląc dostępne w zbiorze danych sceny na 4 podzbiory. Zawsze jeden podzbiór był testowym, a 3 pozostałe treningowymi. Prawie wszystkie sceny w zbiorze danych zawierają palety o jednym kolorze. W każdym z czterech podzbiorów walidacji krzyżowej rozkład scen zawierających palety trzech kolorów był równy.

4.2.1. Detekcja w oparciu o barwę

Klasyfikator zastosowany w detekcji w oparciu o barwę jest Naiwnym klasyfikatorem bayesowskim z rozkładem Gaussa. Jego zadaniem jest klasyfikacja pojedynczych pikseli zareprezentowanych w przestrzeniach barw CMYK, HSV i Lab, z wyłączeniem kanałów kontrastu (K, V i L). Z punktu widzenia klasyfikatora piksel jest więc wektorem liczb całkowitych o długości 7. Dwie klasy, na które dzielimy piksele to "paleta" i "tło". Rezultat klasyfikacji wykorzystywany jest do utworzenia obrazu binarnego, co zostało w pełni opisane w sekcji . Jak opisano w tej samej sekcji, zbiór treningowy i testowy potrzebny do nauczenia oraz testowania jakości klasyfikacji pikseli pozyskiwany jest z obrazów w zbiorze danych poprzez wycięcie oetykietowanych palet, odrzucenie pikseli należących do dziur na widły oraz utworzenie listy pikseli należących do wszystkich palet ze wszystkich ujęć, ale osobnej dla każdego koloru palety w zbiorze danych. W opracowanym w ramach pracy zbiorze danych są to kolor naturalnego drewna, ciemniejszy kolor bejcowanego drewna oraz niebieski. Pozostałe piksele nienależące do palet uznawane są za piksele tła i również zostają skonkatenowane w jedną listę, która jednak ze względu na duży rozmiar może być próbkowana. Wyniki klasyfikacji na tak przygotowanym zbiorze pikseli przedstawiono w tabelach 4.1 oraz 4.2. Wyniki dla koloru niebieskiego odstają wyraźnie od pozostałych kolorów. Palety niebieskie jako jedyne zostały nagrane w ciemnych warunkach oświetleniowych. Niski wynik jest najprawdopodobniej spowodowany charakterystycznym, silnym szumem powstającym w niekorzystnych warunkach. Problem ten został opisany szerzej w sekcji 2.2 dotyczącej przeglądu metod odszumiania. Wyniki dla pozostałych kolorów są zadowalające. Zwłaszcza biorąc pod uwagę, że piksele tła pochodzą w większości przypadków z supermarketu i zawierały różnokolorowe produkty, a ponadto niektóre palety były w pewnym stopniu przesłonięte folią typu "stretch". Wynik ewaluacji całej metody mierzony indeksem Jaccarda wyniósł 0.35.

kolor/miara	accuracy	precision	recall	F1
naturalne drewno	0.888	0.795	0.902	0.844
ciemne drewno	0.898	0.834	0.870	0.850
niebieski	0.851	0.767	0.818	0.790

Tabela 4.1. Tabela z wynikami Naiwnego klasyfikatora bayesowskiego uczonego oraz testowanego na zbiorze pikseli należących do palet o danym kolorze oraz tła.

etykieta		naturalne drewno		ciemne drewno		niebieski	
		paleta		paleta		paleta	
		paleta	tło	tło	paleta	tło	tło
	naturalne drewno	20'207'466	2'361'136	2'289'732	215'571	31'539'136	2'709'726
	tło	5'447'603	39'543'560	394'511	4'616'095	10'149'640	64'487'517
		paleta	tło	paleta	tło	paleta	tło
						predykcja	

Tabela 4.2. Macierze pomyłek dla Naiwnego klasyfikatora bayesowskiego uczonego oraz testowanego na zbiorze pikseli należących do palet o danym kolorze oraz tła.

4.2.2. Detekcja w oparciu o gradient

Klasyfikator zastosowany w detekcji w oparciu o gradient jest Lasem losowym. Jego zadaniem jest klasyfikacja wektorów zawierających fragment obrazu odpowiadający wielkością oknu przesuwnemu, poddany ekstrakcji cech takich jak skala szarości i wielkość gradientu w różnych orientacjach. Cechy opisano szczegółowo w sekcji 3.4. Wejściowy wektor liczb całkowitych ma długość 21'120. Jest to rezultat spłaszczenia 8 okien cech o wymiarach 120x22. Dwie klasy, na które dzielimy piksele to "paleta" i "tło". Rezultat klasyfikacji wykorzystywany jest do oceny prawdopodobieństwa występowania palety w danym oknie. Jak opisano w sekcji 3.4, zbiór danych uczących oraz testowych tworzony jest poprzez wycięcie ze wszystkich obrazów prostokątów o rozmiarach 120x22 px odpowiadających oknu przesuwnemu. Tak samo jak w metodzie opartej o barwę, w wyciętym prostokącie mieści się wykrywana paleta oraz 10 procent marginesu wokół niej (rysunek 3.6). Należy w ten sposób wyciąć ze zbioru danych wszystkie palety oraz co najmniej drugie tyle obszarów zawierających tło. Wyniki klasyfikacji przedstawiono w tabelach 4.3 oraz 4.4. Natomiast wynik ewaluacji całej metody, czyli w połączeniu z techniką okna przesuwnego, mierzony indeksem Jaccarda wyniósł 0.68, zatem o 0.33 więcej niż metody opartej o barwę.

miara	accuracy	precision	recall	F1
wynik	0.981	0.998	0.930	0.960

Tabela 4.3. Tabela z wynikami klasyfikatora Lasu losowego uczonego oraz testowanego na zbiorze prostokątów zawierających palety oraz tło.

etykieta	paleta	2419		177
		3	7254	tło
	paleta	paleta	tło	
				predykcja

Tabela 4.4. Macierz pomyłek dla klasyfikatora Lasu losowego uczonego oraz testowanego na zbiorze prostokątnych okien zawierających palety oraz tło.

4.2.3. Analiza predykcji metod składowych

Podstawowe założenie zespołów klasyfikatorów polega na tym, że klasyfikatory składowe dokonują błędnych predykcji w innych miejscach w dziedzinie. Przykładowo, gdy wykorzystujemy 3 klasyfikatory, liczymy na to, że na wzorcu na którym pomylił się pierwszy z nich, dwa następne dokonają poprawnej predykcji. Wtedy decyzją większościową zespół zaklasyfikuje wzorzec poprawnie. Metod łączenia zespołów klasyfikatorów nie można bezpośrednio przenieść na zespół metod zaproponowany w ramach niniejszej pracy, jednak idea pozostaje ta sama. Obliczono procent predykcji, które były różne dla obu metod. Na 2591 predykcji okien zawierających palety zaledwie 215 było takich samych dla obu metod, czyli 9%. To oznacza, że dla pozostałych 81% palet w zbiorze, jedna z metod poradziła sobie lepiej, a druga gorzej. Niejednomysłość obu metod jest bardzo wysoka, a tym samym potencjalny obszar do poprawy skuteczności - duży.

4.3. BADANIA ZESPOŁU METOD

Po ewaluacji metod bazowych, w oparciu o barwę i gradient, zbadano skuteczność ich połączenia. Zgodnie z przedstawionym w poprzednim rozdziale rysunkiem 3.1, łączenie odbywa się

poprzez agregację dwóch wartości prawdopodobieństw występowania palety w oknie przesuwnym. Dopiero na podstawie zagregowanych prawdopodobieństw odbywa się ostateczny wybór okien zawierających paletę. Przebadano metody agregacji: suma, maksimum, minimum oraz suma ważona. Dokonano także standaryzacji na wartościach prawdopodobieństw, po czym powtórzono badania dla wszystkich czterech metod. Wartości średnich i odchyлеń standardowych obliczonych w celu standaryzacji przedstawiono w tabeli 4.5. Wagi przyjęte w agregacji sumą ważoną to wartości indeksu Jaccarda dla metod składowych. Dla metody opartej o barwę jest to 0.350, a gradient 0.683. Szczegółowy opis metod agregacji i standaryzacji znajduje się w sekcji 3.5. Wyniki przedstawiono w tabeli 4.6. Z obu metod bazowych, metoda oparta o gradient osiągnęła znacznie wyższy wynik. Cztery agregacje okazały się być lepsze od metody bazowej. Są to kolejno suma, minimum ze standaryzacją, suma ważona ze standaryzacją i suma ze standaryzacją. Standaryzacja pozwoliła na poprawę skuteczności dwóch agregacji, to jest minimum i sumy ważonej. Najlepszy wynik osiągnęła najbardziej intuicyjna metoda, czyli suma. Wynik tak połączonego zespołu metod jest o 6,3% lepszy od metody bazowej opartej o gradient i o 107,4% lepszy od metody opartej o barwę. 6,3% to bardzo duża wartość, biorąc pod uwagę, że normy stosowane w przemyśle sięgają 98% i nierzadko twórcy systemów starają się poprawić metody o choćby dziesiątą część procenta. Taka poprawa została osiągnięta poprzez rozszerzenie detekcji na podstawie obrazów w skali szarości o dodanie wnioskowania na podstawie barwy. Należy jednak взять pod uwagę, że wymagało to pogrupowania palet w zbiorze według kolorów oraz zwiększenia złożoności metody. Przed wdrożeniem należało zastanowić się czy uzyskany wzrost skuteczności na pewno rekompensuje wspomniane koszty i wymagania związane z wykorzystaniem barwy.

metoda/miara	średnia	odchylenie st.
barwa	0.460	0.089
gradient	0.047	0.096

Tabela 4.5. Tabela z wartościami średnich i odchyłeń standardowych prawdopodobieństw występowania palety w oknach przesuwnych w autorskim zbiorze danych wyznaczanych metodami opartymi na barwie i gradiencie.

agregacja	indeks Jaccarda
tylko barwa	0.350
tylko gradient	0.683
maksimum	0.629
minimum	0.663
suma	0.726
suma ważona	0.636
maksimum, stand.	0.548
minimum, stand.	0.701
suma, stand.	0.686
suma ważona, stand.	0.688

porządek według kolejności badań

agregacja	indeks Jaccarda
suma	0.726
minimum, stand.	0.701
suma ważona, stand.	0.688
suma, stand.	0.686
tylko gradient	0.683
minimum	0.663
suma ważona	0.636
maksimum	0.629
maksimum, stand.	0.548
tylko barwa	0.350

porządek według indeksu Jaccarda malejąco

Tabela 4.6. Tabela z wynikami ewaluacji metod bazowych i ich połączenia badanymi sposobami agregacji.

4.3.1. Porównanie do istniejących metod

W celu miarodajnego porównania skuteczności opracowanego w ramach niniejszej pracy zespołu metod i metod poznanych w przeglądzie literaturowym, należało zbadać je na tym

samym zbiorze oraz tą samą miarą. Jak wynika z rozważań w sekcji 2.1.4 przeglądu literaturowego, każdy autor wykorzystał inny zbiór danych i inną miarę do ewaluacji swojej metody. Reprezentatywne zbiory danych wykorzystane w poznanych publikacjach nie zostały udostępnione. Najprawdopodobniej w celu uniemożliwienia łatwego odtworzenia metod, ze względu na przemysłowo-biznesowy charakter badań. Jedyną możliwością jest porównanie wszystkich metod na zbiorze danych opracowanym w ramach niniejszej pracy. To z kolei wymaga dostępu do gotowego kodu lub jego reimplementacji na podstawie publikacji. Najprostsze metody można zreimplementować, jednak ich porównywanie mija się z celem, ponieważ przegrywają już na etapie analizy funkcjonalności, np. metody wymagające ręcznego wyznaczania prógu wartości kanałów koloru w celu binaryzacji. Z kolei metody bardziej złożone i funkcjonalne jak Cucciara et al.[12], Syu et al. [40] czy Varga et al. [44][46] nie są dostępne w sieci, a ich implementacja byłaby zadaniem zbyt czasochłonnym na końcowy etap pisania niniejszej pracy, o ile w ogóle możliwym, ze względu na częstą praktykę ukrywania niezbędnych szczegółów i parametrów przez autorów.

Można jedynie udowodnić wyższą skuteczność opracowanej metody względem zreimplementowanej pracy Varga et al. z roku 2015 opartej o gradient, która stanowi metodę składową zaproponowanego rozwiązania. Pozostałe prace tych samych autorów z lat 2014 i 2016 [44][46] osiągały w badaniach twórców bardzo zbliżoną skuteczność zgodnie z tabelą 2.2, stąd można spodziewać się, że również wypadłyby słabiej.

4.3.2. Podsumowanie

Przeprowadzono badania metod składowych oraz zespołu metod. Wszystkie testy zostały wykonane na tym samym autorskim zbiorze oraz według jednej uniwersalnej miary, czyli indeksu Jaccarda. Zastosowano walidację krzyżową z podziałem na podzbiory scen, dzięki czemu otrzymane wyniki są obiektywne i pełne. Uzyskane rezultaty pokazały, że metoda oparta o gradient, zapożyczona z pracy Varga et al. z roku 2015 [45] wykazuje wysoką skuteczność. Zgodnie z oczekiwaniemi rozszerzenie jej o detekcję na podstawie barwy, pozwoliło na znaczny wzrost skuteczności o 6,3%. Niewątpliwie osiągnięty sukces jest jednak obarczony istotnymi wymaganiami. Pierwszym z nich jest konieczność obecności w zbiorze danych palet w kolorach, które chcemy wykrywać, a także dodanie informacji o kolorze na etapie etykietowania. Jednak stworzenie większego zbioru, posiadającego większość stosowanych typów palet nie stanowiłoby prawdopodobnie większego problemu dla producenta dysponującego podstawowymi środkami finansowymi. Drugim jest nieunikniony, nieco większy nakład obliczeniowy. Niniejsza praca badawcza zupełnie nie poruszała ważnego problemu wydajności, jednak wdrożenie zaproponowanej metody wymagałoby sprawdzenia szybkości działania na efektywnej implementacji. Niemniej wykorzystane Naiwne klasyfikatory bayesowskie należą do skrajnie lekkich i szybkich w działaniu, więc nie powinny przysparzać problemów. Opracowano więc metodę, która może zostać z powodzeniem wykorzystana w rzeczywistości.

ZAKOŃCZENIE

W ramach pracy udało się opracować metodę detekcji palet w sekwencji wideo pochodzącej z kamery zamontowanej na autonomicznym wózku widłowym. Zaproponowane rozwiązanie osiągnęło skuteczność wyższą o 6,3% od jednej z najlepszych istniejących metod. Przed przystąpieniem do projektowania, dokonano wyczerpującego przeglądu literaturowego zakończonego kilkoma kluczowymi tabelami w celu ustrukturyzowania pozyskanej wiedzy. Z tak przygotowanych danych dotyczących dotychczasowych publikacji wyciągnięto wnioski, które pozwoliły na wskazanie braków i ukierunkowały projekt rozwiązania oraz badania w dalszej części pracy.

Wyróżniono trzy obszary wymagające badań. Odszumianie obrazów, wygładzanie detekcji po czasie oraz integrację metod pracujących na obrazach w skali szarości i kolorowych. Ostatnie zagadnienie zostało zaadresowane w pracy poprzez zaprojektowanie i zbadanie autorskiego zespołu metod. Pierwsza z metod składowych, oparta o barwę, została skonstruowana przez autora. W celu uniknięcia wad dotychczasowych rozwiązań, z których główna to ręczne wyznaczanie progów wartości kanałów koloru w celu binaryzacji, zaproponowano wykorzystanie Naiwnych klasyfikatorów bayesowskich dla każdego koloru rozpoznawanych typów palet. Taki zabieg pozwolił na automatyczne nauczenie klasyfikatora detekcji pikseli należących do palety, bez działania ręcznego. Z kolei druga metoda składowa, oparta na gradiencie pozyskiwanym z obrazów w skali szarości, została zapożyczona z pracy Varga et al. z roku 2015 [45]. Obie metody składowe połączono na etapie obliczania prawdopodobieństwa palety w oknie przesuwnym, poprzez agregację wartości otrzymywanych z obu metod. Zbadano kilka funkcji agregacji, z których suma pozwoliła na uzyskanie najwyższego wyniku, przewyższającego rezultaty metod bazowych testowanych osobno.

Wszystkie testy przeprowadzono na zbiorze danych zebranym specjalnie do badań w ramach niniejszej pracy. Zbiór zawierał 29 scen najazdów wózka na wprost palety, łącznie ponad 2250 obrazów. Zastosowano uniwersalną i spopularyzowaną miarę nazywaną indeksem Jaccarda, która uniezależniła badania od mnogich, specyficznych metod proponowanych w dotychczasowych publikacjach. W celu uzyskania wyniku niezależnego od podziału danych, zastosowano walidację krzyżową z podziałem na cztery podzbiory scen.

Sukces opracowanej metody został osiągnięty kosztem dwóch głównych wymogów. Pierwszy to konieczność występowania w zbiorze danych palet o kolorze, który chcemy wykrywać, a także dodania informacji o barwie do każdej ręcznie wyznaczanej etykiety. Druga polega na zwiększeniu złożoności obliczeniowej poprzez osobne przetwarzanie obrazu w skali szarości i kolorowego. Wdrożenie rozwiązania wymagałoby zbudowania większego zbioru danych oraz przetestowania efektywności implementacji. Badania przeprowadzone na tym etapie udowadniają, że autorska metoda ma wielki potencjał w rzeczywistych zastosowaniach.

BIBLIOGRAFIA

- [1] Aharon, M., Elad, M., Bruckstein, A., *K-svd: An algorithm for designing overcomplete dictionaries for sparse representation*, IEEE Transactions on Signal Processing. 2006, tom 54, 11.
- [2] Allen, J., Xu, R., *Tracking using camshift algorithm and multiple quantized feature spaces*. (2003), str. 3–7.
- [3] Aref, M.M., Ghabcheloo, R., Mattila, J., *A macro-micro controller for pallet picking by an articulated-frame-steering hydraulic mobile machine*, w: 2014 IEEE International Conference on Robotics and Automation (ICRA) (2014).
- [4] Beck, A., Teboulle, M., *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Transactions on Image Processing. 2009, tom 18, 11.
- [5] Benesty, J., Chen, J., Huang, Y., *Study of the widely linear wiener filter for noise reduction*, w: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (2010).
- [6] Buades, A., Coll, B., Morel, J., *A non-local algorithm for image denoising*, w: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), tom 2 (2005).
- [7] Byun, S., Kim, M., *Real-time positioning and orienting of pallets based on monocular vision*, w: 2008 20th IEEE International Conference on Tools with Artificial Intelligence, tom 2 (2008).
- [8] Chen, C.Y., Hsia, C.H., Yang, C.Y., *Evolutionary design of multiplierless lifting-based 2d dwt filters for low-resolution image processing*, Multimedia Tools and Applications. 2015, tom 75.
- [9] Chen, G., Peng, R., Wang, Z., Zhao, W., *Pallet recognition and localization method for vision guided forklift*, w: 2012 8th International Conference on Wireless Communications, Networking and Mobile Computing (2012).
- [10] Chen, Y., Pock, T., *Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration*. 2015.
- [11] Cucchiara, R., Filicori, F., *A highly selective ht based algorithm for detecting extended, almost rectilinear shapes*, w: Computer Analysis of Images and Patterns, pod red. V. Hlaváč, R. Šára (Springer Berlin Heidelberg, Berlin, Heidelberg, 1995).
- [12] Cucchiara, R., Piccardi, M., Prati, A., *Focus based feature extraction for pallets recognition* (2000).
- [13] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K., *Image denoising by sparse 3-d transform-domain collaborative filtering*, IEEE transactions on image processing : a publication of the IEEE Signal Processing Society. 2007, tom 16.
- [14] Dalal, N., Triggs, B., *Histograms of oriented gradients for human detection*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005). 2005, tom 2.
- [15] Dollar, P., Appel, R., Belongie, S., Perona, P., *Fast feature pyramids for object detection*, Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2014, tom 36.
- [16] Dong, W., Zhang, L., Shi, G., Li, X., *Nonlocally centralized sparse representation for image restoration*, IEEE Transactions on Image Processing. 2013, tom 22, 4.
- [17] Fan, L., Zhang, F., Fan, H., Zhang, C., *Brief review of image denoising techniques*, Visual Computing for Industry, Biomedicine, and Art. 2019, tom 2.
- [18] Gajdzica, M., *Filtr kalmana od teorii do praktyki - #1 - Matlab*, <https://forbot.pl/blog/filtr-kalmana-teorii-praktyki-1-id2855>. Ostatni dostęp: 2020-03-14.
- [19] Garibotti, G., Masciangelo, S., Ilic, M., Bassino, P., *Robolift: a vision guided autonomous fork-lift for pallet handling*, w: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS '96, tom 2 (1996), str. vol.2.
- [20] Garibotto, G., Masciangelo, S., Ilic, M., Bassino, P., *Service robotics in logistic automation: Robolift*:

- vision based autonomous navigation of a conventional fork-lift for pallet handling*, w: *1997 8th International Conference on Advanced Robotics. Proceedings. ICAR'97* (1997).
- [21] Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X., Zhang, L., *Weighted nuclear norm minimization and its applications to low level vision*, International Journal of Computer Vision. 2016, tom 121.
 - [22] Guang-zhao Cui, Lin-sha Lu, Zhen-dong He, Li-na Yao, Cun-xiang Yang, Bu-yi Huang, Zhi-hong Hu, *A robust autonomous mobile forklift pallet recognition*, w: *2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010)*, tom 3 (2010).
 - [23] Haralick, R., Shapiro, L., *Computer and Robot Vision*, nr t. 2 w Computer and Robot Vision (Addison-Wesley Publishing Company, 1993).
 - [24] He, D.C., Wang, L., *Texture unit, texture spectrum, and texture analysis* (1990), str. 509–512.
 - [25] Kalman, R.E., *A new approach to linear filtering and prediction problems*, Transactions of the ASME—Journal of Basic Engineering. 1960, tom 82, Series D.
 - [26] Kim, W., Helmick, D., *Model-based object pose refinement for terrestrial and space autonomy*. 2000.
 - [27] Liu, X., Tao, D., Song, M., Zhang, L., Bu, J., Chen, C., *Learning to track multiple targets*, IEEE transactions on neural networks and learning systems. 2014, tom 26.
 - [28] O’Gorman, F., Clowes, M., *Finding picture edges through collinearity of feature points.*, IEEE Trans. Computers. 1976, tom 25.
 - [29] Pagès, J., Armangué, X., Salvi, J., Freixenet, J., Martí, J., *A computer vision system for autonomous forklift vehicles in industrial environments I*.
 - [30] Penoyer, R., *The alpha-beta filter*, C User’s Journal. 1993, tom 11, 7.
 - [31] Rajamani, M.R., *Data-based techniques to improve state estimation in model predictive control* (2007).
 - [32] RC, G., RE, W., *Digital image processin*, 3rd edn. (PrenticeHall, Inc, Upper Saddle River, 2006).
 - [33] Rudin, L.I., Osher, S., Fatemi, E., *Nonlinear total variation based noise removal algorithms*, w: *Proceedings of the Eleventh Annual International Conference of the Center for Nonlinear Studies on Experimental Mathematics: Computational Issues in Nonlinear Science: Computational Issues in Nonlinear Science* (Elsevier North-Holland, Inc., USA, 1992).
 - [34] Ruikang Yang, Lin Yin, Gabbouj, M., Astola, J., Neuvo, Y., *Optimal weighted median filtering under structural constraints*, IEEE Transactions on Signal Processing. 1995, tom 43, 3.
 - [35] Sabattini, L., Aikio, M., Beinschob, P., Boehning, M., Cardarelli, E., Digani, V., Krengel, A., Magnani, M., Mandici, S., Oleari, F., Reinke, C., Ronzoni, D., Stimming, C., Varga, R., Vatavu, A., Lopez, S., Fantuzzi, C., Mayra, A., Nedevschi, S., Fuerstenberg, K., *The pan-robots project: Advanced automated guided vehicle systems for industrial logistics*, IEEE Robotics Automation Magazine. 2017, tom PP.
 - [36] Sas, J., Paradowski, M., *Analiza obrazów i wideo*, 2020, wykład 13, Analiza wideo - predykcja ruchu, https://eportal.ii.pwr.edu.pl/w08/BOARD/Sas.Jerzy/Analiza%20obraz%C3%B3w%20i%20wideo%20-%202019%202020/Wyk%C5%82ady/A0iW_13.pdf. Ostatni dostep: 2020-03-14.
 - [37] Seelinger, M., Yoder, J., *Automatic pallet engagment by a vision guided forklift*, w: *Proceedings of the 2005 IEEE International Conference on Robotics and Automation* (2005).
 - [38] Smith, S.M., Brady, M., *Susan—a new approach to low level image processing*, International Journal of Computer Vision. 1997, tom 23.
 - [39] Sutour, C., Deledalle, C., Aujol, J., *Adaptive regularization of the nl-means: Application to image and video denoising*, IEEE Transactions on Image Processing. 2014, tom 23, 8.
 - [40] Syu, J.L., Li, H.T., Chiang, J.S., Hsia, C.H., Wu, P.H., Hsieh, C.F., Li, S.A., *A computer vision assisted system for autonomous forklift vehicles in real factory environment*, Multimedia Tools and Applications. 2016, tom 76.
 - [41] Tikhonov, A., Arsenin, V., *Solutions of ill-posed problems*, Scripta series in mathematics (Winston, 1977).

- [42] Tomasi, C., Manduchi, R., *Bilateral filtering for gray and color images*, w: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)* (1998), str. 839–846.
- [43] Trucco, E., Verri, A., *Introductory techniques for 3-D computer vision*. (1998).
- [44] Varga, R., Nedevschi, S., *Vision-based autonomous load handling for automated guided vehicles*, w: *2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP)* (2014).
- [45] Varga, R., Costea, A., Nedevschi, S., *Improved autonomous load handling with stereo cameras* (2015).
- [46] Varga, R., Nedevschi, S., *Robust pallet detection for automated logistics operations* (2016).
- [47] Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A., *Extracting and composing robust features with denoising autoencoders*, w: *Proceedings of the 25th International Conference on Machine Learning, ICML '08* (Association for Computing Machinery, New York, NY, USA, 2008).
- [48] Viola, P., Jones, M., *Robust real-time face detection*, International Journal of Computer Vision. 2004, tom 57.
- [49] Xie, J., Xu, L., Chen, E., *Image denoising and inpainting with deep neural networks*, Advances in Neural Information Processing Systems. 2012, tom 1.
- [50] Xu, S., Cornell, A.C., *Robust object tracking using kalman filters with dynamic covariance* (2011).
- [51] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L., *Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising*, IEEE Transactions on Image Processing. 2017, tom 26.
- [52] Zhang, K., Zuo, W., Zhang, L., *Ffdnet: Toward a fast and flexible solution for cnn based image denoising*, IEEE Transactions on Image Processing. 2017, tom PP.