# POLITECHNIKA WROCŁAWSKA

WYDZIAŁ ELEKTRONIKI

KIERUNEK: Automatyka i Robotyka

SPECJALNOŚĆ: Technologie informacyjne w systemach automatyki

PRACA DYPLOMOWA

MAGISTERSKA

Badanie metod rozpoznawania znaków drogowych z wykorzystaniem wielu źródeł obrazu  
Research on traffic sign recognition methods using multiple image source

autor : BARTOSZ LENARTOWICZ

autor : BARTOSZ LENARTOWICZ

Opiekun pracy:

Dr inż. Bartosz Jabłoński W04/K8

OCENA PRACY:

WROCŁAW 2019

# Streszczenie

[here] (streszczenie na sam koniec)

**Spis treści**

[Streszczenie 2](#_Toc9261981)

[Analiza problemu, cel i zakres pracy 4](#_Toc9261982)

[Metody detekcji obiektów na obrazach 6](#_Toc9261983)

[Zastosowania różnych typów rejestratorów obrazu 8](#_Toc9261984)

[Tworzenie panoramy na potrzeby systemu TSR 8](#_Toc9261985)

[Metody bazujące na punktach charakterystycznych 10](#_Toc9261986)

[Podsumowanie tworzenia panoramy na potrzeby systemu TSR 12](#_Toc9261987)

[Segmentacja obrazów 12](#_Toc9261988)

[Wykrywanie znaków na podstawie koloru 13](#_Toc9261989)

[Wykrywanie na podstawie kształtu 16](#_Toc9261990)

[Wykrywanie na podstawie tekstury 19](#_Toc9261991)

[Podejście hybrydowe oraz inne podejścia 20](#_Toc9261992)

[Klasyfikacja znaków drogowych 26](#_Toc9261993)

[Wzorcowe dopasowanie 26](#_Toc9261994)

[Sztuczne sieci neuronowe 26](#_Toc9261995)

[Maszyna wektorów nośnych do klasyfikacji 30](#_Toc9261996)

[Drzewo k-wymiarowe 31](#_Toc9261997)

[Lasy losowe 31](#_Toc9261998)

[Podsumowanie 32](#_Toc9261999)

[Program 34](#_Toc9262000)

[Propozycja rozwinięcia/konstrukcji metody 35](#_Toc9262001)

[Przygotowanie środowiska badawczego i plan badań 36](#_Toc9262002)

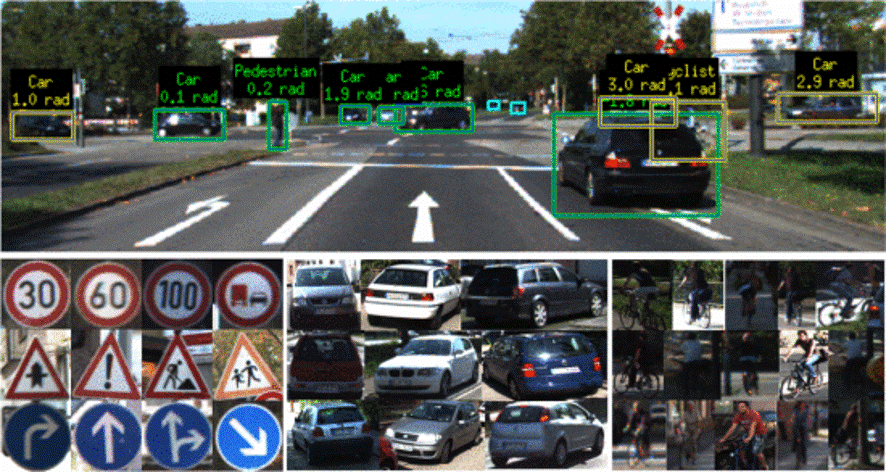
[Podsumowanie 37](#_Toc9262003)

[Bibliografia 38](#_Toc9262004)

# Analiza problemu, cel i zakres pracy

Automatyczne rozpoznanie znaków drogowych jest aspektem niosącym ze sobą wiele korzyści, szczególnie jeżeli ich analiza następuje w czasie rzeczywistym. Korzyścią może być wspomaganie kierowców podczas jazdy. Taki system, nazywany ADAS (ang. Advanced Driver Assistance System) zostaje implementowany w wielu współczesnych samochodach i poprawia komfort i bezpieczeństwo prowadzenia samochodu. Po nieco dłuższym zastanowieniu można wywnioskować, że rozpoznawanie znaków również jest niezbędne dla autonomicznych pojazdów. Muszą one mieć możliwość adaptacji do zmieniającej się infrastruktury drogi oraz zasad na niej panującej np. w trakcie robót budowlanych. Oprócz dwóch powyższych aspektów rozpoznanie znaków drogowych może realnie przyczynić się do polepszania bezpieczeństwa na drodze. Badając w jakim stopniu znak jest czytelny np. czy nie został zamazany, zakrzywiony system może informować deformacjach służbę drogową. Zastosowanie automatycznego rozpoznawania znaków znajduje zastosowanie w coraz to nowych dziedzinach z całą pewnością można rzec, że taki aspekt będzie poruszany w wielu pracach badawczych. Warto wspomnieć, że dziedzina jest rozwijana od ponad dwóch dekad. Jedna z pierwszych prac nad systemem została przedstawiona w roku 1996 w pracy [1].

Starając się wyjaśnić zakres pracy magisterskiej należy rozszerzyć kilka terminów, które zostały w niej poruszone. W literaturze problem rozpoznawania znaków drogowych przedstawia się jako problem TSR (ang. Traffic Sign Recognition). Jest on składową większego zagadnienia jakim jest analiza sytuacji w okuł samochodu podczas jazdy. W takim przypadku obszar wokół samochodu traktowany jest jako scena, w której należy rozpoznawać i analizować obiekty na niej umieszczone. Zostaje to wspomniane, ponieważ niektóre systemy prócz rozpoznania znaków drogowych są w stanie rozpoznawać inne obiekty takie jak samochody, rowerzystów czy pieszych [2].



Rysunek Rodzaje obiektów na i wokół drogi

Rozpoznanie większej ilości obiektów rozszerza możliwości systemu i umożliwia śledzenie obiektów ruchomych np. w celu poinformowania kierowców o kursie kolizyjnym. Śledzenie pozwala również zapobiec wykrywaniu kilkukrotnie tego samego obiektu. Według autorów pracy [3] aspekt śledzenia jest szczególnie ważny dla systemów, w których pojazd jest kierowany przez kierowcę (a nie przez algorytm). Wielokrotne informowanie kierowcy o znaku może rozproszyć jego uwagę powodując tym samym zagrożenie. Autorzy w swojej pracy również proponują traktowanie kierowcy jako integralną część środowiska i śledzenie jego ruchów dla lepszej integracji systemu z kierowcą.

Niniejsza praca została stworzona z myślą o wzbogaceniu systemu wykrywania znaków drogowych przez podłączenie do systemu kolejnych kamer. W celu realizacji pracy dyplomowej należało wykonać kilka zadań. Pierwszym z nich był przegląd literaturowy metod lokalizacji obiektów na obrazie oraz metod rozpoznawania znaków drogowych. Następnie należało przeanalizować możliwości wykorzystania wielu źródeł obrazu. Kluczowym etapem było opracowanie algorytmu detekcji znaków na podstawie więcej niż jednego źródła obrazu. Po jego opracowaniu należało zaimplementować algorytm w wybranym środowisku oraz przygotować dane, które mogły by posłużyć do testów. Najważniejszym etapem dla pracy było wykonanie eksperymentów i przeanalizowanie ich wyników. Na sam koniec należało przedstawić uzyskane wyniki czytelnikowi i zredagować całą pracę w taki sposób by była spójna i czytelna.

Opis pracy dyplomowej został stworzony w następujący sposób. Wpierw zostało przedstawione streszczenie zawierające krótki opis pracy. Na następnej stronie został umieszczony spis treści. Następnie opisano wstęp pracy, w którym przedstawiono tematykę problemu, zadania do wykonania i strukturę pracy. …

# Metody detekcji obiektów na obrazach

W niniejszym rozdziale zostały przedstawione metody, które starają się sprostać zadaniu rozpoznawania obiektów na obrazie. Większość z przedstawionych metod skupia się na wykrywaniu i rozpoznaniu pionowych znaków drogowych. Zazwyczaj przedstawiane metody są podzielone na dwa etapy. Pierwszym z nich jest wyszukanie na obrazie regionów mogących zawierać znak drogowy lub obiektów przypominający taki znak. Taki etap nazwany jest segmentacją obrazu. Ponieważ znaki drogowe mają konkretnie określone kolory i kształty, w przypadku problemu TSR etap segmentacji prawie zawsze poprawia szybkość algorytmu. Metody segmentacji zostały przedstawione w podrozdziale Segmentacja obrazów. Dla rejonów wyselekcjonowanych przez algorytmy segmentacji należy sprawdzić czy znak drogowy faktycznie na nich istnieje i jeżeli istnieje rozpoznać jakie ma on znaczenie. Etap ten nazywany jest klasyfikacją, a metody odpowiadające za klasyfikacje zostały przedstawione w rozdziale Klasyfikacja znaków drogowych.

Przedstawiając metody odpowiedzialne za segmentację i klasyfikację należy zaznaczyć, że znaki drogowe nie zostały ujednolicone na całym świecie. Istnieją dwie główne konwencje starające się je ujednolicić. Pierwszą z nich jest grupa znaków opartych według zasad MUTCD (ang. Manual on Uniform Traffic Control Devices). Tworzenie znaków według tych restrykcji obowiązuje głównie w krajach położonych w Ameryce Północnej i Południowej, Australii i krajach azjatyckich położnych blisko Australii. Kraje położone w Europie, Azji oraz Afryce używają znaków drogowych zaprojektowanych według Konwencji Wiedeńskiej. Oczywiście istnieją kraje np. Argentyna, Brazylia, Chiny, które przyjmują część zasad z jednej, a część z drugiej konwencji lub projektują je według własnych zasad. Znaki każdej konwencji mają ujednolicone wzory tworzenia i konkretnie przypisane kolory i kształty dla każdego typu znaku. Konwencja również wprowadziła zasady obowiązujące dla znaków poziomych i sygnalizacji świetlnej. Niestety dopuszcza ona rozbieżności według projektowania znaków. W większości krajów znaki ostrzegawcze Konwencji Wiedeńskiej są przedstawiane jako biały równoboczny trójkąt z czarnym symbolem pośrodku czerwoną obwódką, lecz w kilku krajach w tym w Polsce białe tło zastąpiono kolorem żółtym. Problem różnorodności znaków został to wspomniany, ponieważ wszystkie rozwiązania problemu TSR, w tym te przedstawione w niniejszej pracy, bazują na wykrywaniu znaków za pomocą algorytmów bazujących na kolorze czy kształcie. Opisywane prace skupiać się będą wokół wykrywania znaków tworzonych według Konwencji Wiedeńskiej.

Ważnym aspektem każdej pracy baza obrazów ze znakami drogowymi. Niekiedy w starszych algorytmach autorzy samodzielnie selekcjonowali klatki filmu, zaś większość nowych publikacji zostało przetestowanych na istniejących bazach danych. Powszechnie wykorzystywane publiczne bazy danych to baza:

* [German Traffic Sign Recognition Benchmark](http://benchmark.ini.rub.de/?section=gtsrb&subsection=news),
* [BelgiumTS Dataset](https://btsd.ethz.ch/shareddata/),
* [Chinese Traffic Sign Database](http://www.nlpr.ia.ac.cn/pal/trafficdata/recognition.html)
* [Linkopings Universitet Traffic Signs Dataset](http://www.cvl.isy.liu.se/research/datasets/traffic-signs-dataset/download/),
* [RUG Traffic Sign Image Database](http://www.cs.rug.nl/~imaging/databases/traffic_sign_database/traffic_sign_database.html).

Zbiorem wyjściowym dla większości publikacji jest pierwszy przedstawiony zbiór danych drogowych GTSRB, w tłumaczeniu Niemiecki Benchmark Znaków Drogowych. Baza ta złożona jest z 43 klas znaków pochodzących z Niemiec. Każda klasa posiada od kilku do kilkunastu osobnych znaków drogowych. Każdy znak został przedstawiony 30 razy obrazach powstałych poprzez wyselekcjonowanie z sekwencji wideo następujących po sobie klatek filmu. Tym sposobem każda klasa posiada od 200 do 2000 obrazów. Obraz jest złożony głównie ze znaku, który wypełnia ok 80 % obrazu. Rysunek 2 przedstawia 30 obrazów jednego znaku z tego zbioru. Drugi ze zbiorów BelgiumTS posiada 62 klasy znaków. Znaki zostały sfotografowane pojedynczo i również zajmują około 80 % obrazu. Każda klasa posiada od 5 do 300 obrazów. Zbiór trzeci CTSD jest to zbiór posiadający jak poprzednio znaki zajmujące 80 % obrazu posiadający 58 klas znaków pochodzących z Chin. Większość znaków chińskich projektowane jest według zasad Konwencji Wiedeńskiej, lecz niektóre z nich tworzone bez konkretnych zasad oraz zamiast symboli posiadają Chińskie znaki. Niemniej jednak część prac zwłaszcza azjatyckich bazuje na tym zbiorze. Zbiór czwarty LUTSD jest Szwedzkim zbiorem posiadający inaczej niż w przypadku poprzednich zbiorów całe obrazy pochodzące z ulicy, gdzie znaki posiada niewielką część obrazów oraz część obrazów bez znaków. Uniwersytet Linkopings dodał adnotacje do 1940 takich obrazów, a około 4 razy tyle udostępnił bez adnotacji. Ostatnim zbiorem jest Rosyjski zbiór danych zawierający 48 obrazów podpisanych w tak jak nazywają się znaki znajdujące się na nich. Gdyby nie różnica w kolorze znaków ostrzegawczych w dwóch ostatnich zbiorach mógłby być przydatne do testowania budowanego algorytmu.

|  |
| --- |
| Obraz zawierający rozmyty, wewnątrz  Opis wygenerowany automatycznieObraz zawierający rozmyty, siedzi, transport  Opis wygenerowany automatycznieObraz zawierający rozmyty, autobus  Opis wygenerowany automatycznie  Obraz zawierający zewnętrzne, drzewo  Opis wygenerowany automatycznieObraz zawierający drzewo  Opis wygenerowany automatycznieObraz zawierający drzewo, znak, zewnętrzne  Opis wygenerowany automatycznie |

Rysunek Znak Ustąp pierwszeństwa ze zbioru GTSRB

Głównym zagadnieniem pracy jest wpływ zastosowania większej ilości kamer dla wykrywania znaków drogowych. Taki problem może być poruszany na kilka sposobów. Pierwszym z nich jest zastosowanie różnych typów rejestratorów obrazu. Skierowane na ten sam obszar w różny sposób odwzorowują otoczenie na obrazie co może powodować polepszenie się skuteczności wyszukiwania zwłaszcza w trudnych warunkach oświetleniowych. Takie zagadnienie zostało przedstawione w akapicie: Zastosowania różnych typów kamer. Inną możliwością jest objęcie przez kilka kamer jednego typu obszaru, który byłby niemożliwy do objęcia przez jedną kamerę. Obrazy uzyskane w ten sposób mogą być analizowane osobno lub można połączyć je ze sobą tworząc obraz panoramiczny, jeżeli w jakimś stopniu się. Algorytmy tworzenia obrazu panoramicznego i wpływ na zagadnienie TSR został przedstawione w akapicie Algorytmy łączenia obrazów w panoramę. Niestety żadna z znalezionych artykułów dla pracy dyplomowej nie rozwija tematu zastosowania większej ilości kamer, dlatego posłużono się publikacjami, które zostały stworzone na potrzeby systemów przeznaczonych do innych zadań.

## Zastosowania różnych typów rejestratorów obrazu

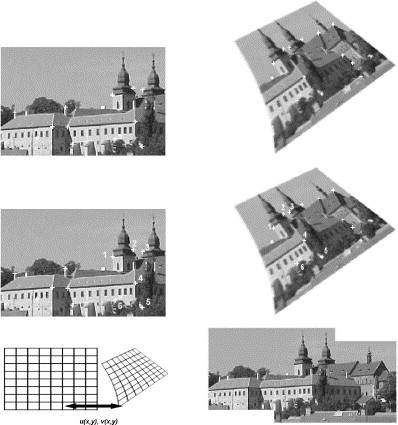
Oprócz klasycznej kamery dostarczającej obraz widzialny przez człowieka istnieje wiele typów rejestratorów obrazu mogących polepszyć wykrywanie znaków w trudnych warunkach. Z wszystkich najbardziej pomocny dla rozpoznawania sytuacji na drodze wydaje się kamera z oświetlaczem podczerwieni. Przetwarzanie obrazu dostarczonego przez obie kamery może odbywać się równolegle lub możliwa jest połączenie obrazów. W tym miejscu można przytoczyć pracę [6], w której został zaproponowany algorytm do łączenia klatek filmowych tego samego obszaru nagrywanych różnymi kamerami. Autorzy testowali algorytm na dwóch kamerach – TV i IR. Ponieważ zazwyczaj kamery rejestrują obraz o różnych rozdzielczościach, algorytm przeskalowuje obrazy do jednej wielkości, a następnie zostają wyszukane krawędzie na obrazie. Autorzy pracy nie podają jaką metodą wyszukują krawędzie. W większości przypadków prac taką metodą jest metoda Canny opisana w akapicie Detektor Canny. Następnie jeden z obrazów zostaje podzielony pionowo na równe części. Każdy z wycinków algorytm stara się dopasować do obrazu drugiego. Najlepsze dopasowanie każdego odcinka wyznacza się za pomocą sumy najmniejszych kwadratów lub sumy modułów różnic. Następnie za pomocą jednej z metod: średniej, mediany lub dominanty wybiera się najlepsze dopasowanie ogólne. Na podstawie dopasowania ogólnego zostaje wyznaczona translacja pozioma obrazów. Warto zaznaczyć, że dominanta uzyskała najlepsze wyniki dopasowania. Gdy obrazy zostają przesunięte o wyznaczoną wartość i proceder powtarza się dla translacji pionowej. Algorytm jest słabo odporny na rotację, jego największą zaletą jest szybkość. Opisując źródła, z których można czerpać inspirację dla fuzji obrazów nie sposób nie wspomnieć o medycynie. Przykładem, który można podać jest tomografia lub rezonans magnetyczny, gdzie nie dość, że pojedyncze klatki obrazów są łączone w taki sposób by tworzyć obraz 3-D to jeszcze często na potrzeby dokładniejszej rozeznania łączy się obrazy 3-D z różnych źródeł. Artykuł [5] przedstawia metodę Powella z użyciem metody Brenta na potrzeby łączenia obrazów medycznych. Chcąc podsumować rozwiązanie zastosowania różnych typów rejestratorów dla problemu TSR można stwierdzić, że jest to dział niosący wiele możliwości, lecz dotychczas niebadany.

## Tworzenie panoramy na potrzeby systemu TSR

W przypadku zarejestrowania otoczenia wokół samochodu przez wiele kamer, gdzie każda kamera obejmuje jego wycinek, jedną z pierwszych rzeczy jaka się nasuwa jest stworzenie obrazu panoramicznego. Taka innowacja może posiadać kilka zalet. Pierwszą z nich jest posiadanie ciągłości obrazu otocznia wokół samochodu. Rozpoznanie znaków na takim obrazie nie powoduje sytuacji konfliktowych w przypadku rozpoznania jednego znaku inaczej na dwóch obrazach. Przykładowo, gdy mamy dwa obrazy niepołączone ze sobą, na jednym obrazie możemy rozpoznać inny znak niż na drugim co wprowadza problem rozstrzygnięcia, który ze znaków jest bardziej prawdopodobny. Jedną z bardziej istotnych zalet systemu z panoramą jest możliwość wychwycenia raz obiektu i jego śledzenie do czasu, kiedy nie zniknie z pola widzenia wszystkich kamer. Jak zostało wspomniane we wstępie jest to bardzo istotne dla systemów ADAS. Dodatkowo przy segmentacji obrazu otrzymanego poprzez panoramę przetwarzane jest mniej jest analiza mniejszej ilości obrazów na etapie segmentacji.

Istnieje szereg algorytmów starających się sprostać problemowi dopasowania kilku obrazów do siebie. Istotną cechą poszukiwanego algorytmu powinna być odporność na różnicę w jasności w obrazach wejściowych oraz szumy spowodowane drganiami i rotacją oraz zabrudzeniami typu: błoto, kurz, krople wody. W wielu pracach np. [4] łączenie obrazów w panoramę zostaje podzielone na kilka etapów. W artykule starano się uniwersalną metodę mogącą łączyć obrazy tej samej sceny uzyskane w różnych czasach, z różnych punktów widzenia lub zarejestrowane różnymi czujnikami. Według pracy proponuje się cztery etapy łączenia obrazów przedstawione poniżej oraz przedstawione na rysunku Rysunek 3:

* Wykrywanie cech kluczowych, gdzie wykrywa się np. kontury, narożniki, centra grawitacji, punkty kluczowe.
* Dopasowywanie cech, gdzie ustalana jest zgodność między cechami wykrytymi w wykrytym obrazie i wykrytymi w obrazie odniesienia. W tym celu używane są różne deskryptory funkcji i podobieństwa oraz relacje przestrzenne między funkcjami.
* Ujednolicenie dwóch modeli. Na tym etapie przekształca się oba obrazy w taki sposób by można było połączyć je w całość za pomocą funkcji odwzorowania. Funkcje odwzorowania umożliwiają wyrównanie obrazów, a nawet zniwelowanie odkształceń. Parametry funkcji odwzorowania są obliczane za pomocą ustalonej zgodności cech.
* Ponowne próbkowanie obrazu i transformacja. Obrazy są transformowane za pomocą funkcji odwzorowania. Wartości obrazu w niecałkowitych współrzędnych są obliczane za pomocą odpowiedniej techniki interpolacji.



Rysunek 3 Przedstawienie 4 kroków tworzenia obrazów

W każdym etapie należy zadecydować, które funkcje będą najskuteczniejsze. Warto pamiętać, by wybrany algorytm wykrywał te same cechy kluczowe na obu obrazach, przy czym każdy z obrazów może być odmiennie zaszumiony. W tej fazie poprzez ekstrakcje cech szczególnych klatki wyłania się punkty kluczowe, niezmiennicze bądź krawędzie obiektów służące do dopasowania obrazów. Po dopasowaniu obrazów następuje fuzja właściwa powodująca scalenie obrazu w panoramę. Warto zaznaczyć, że wyekstraktowane cechy mogą posłużyć w kolejnych etapach do identyfikacji obiektów.

Poniżej zostały omówione różne algorytmy tworzenia panoramy bazujące na punktach kluczowych, charakterystycznych. Są to metody najpopularniejsze mające największe znaczenie w rzeczywistości. Można również wspomnieć o metodzie przedstawionej w artykule [7]. W skrócie metoda polega podzieleniu obu obrazów na niewielkie obszary i dopasowaniu obrazów do siebie za pomocą szybkich, mało dokładnych metod z wykorzystaniem miary statystycznej korelacji. Następnie używając metod bardziej złożonych obliczeniowo, ale dokładniejszych łączy się segmenty w całość i znajduje najlepsze globalne dopasowanie. Autorzy algorytmu w opisie przedstawiają, że algorytm wykorzystuje metodę quasi-Newtona, z piramidą obszarów. Takie łączenie obrazów pozwala na większą eliminację błędów.

### Metody bazujące na punktach charakterystycznych

Metody łączenia obrazów bazujące na wspólnych punktach kluczowych są najbardziej popularnymi metodami znajdującymi zastosowanie w tworzeniu panoramy np. z obrazów zrobionych smartphonem, ponieważ posiadają dużą odporność na zakłócenia. Wyłanianie takich punktów bazuje na ekstrakcji cech obrazu. W literaturze punkty charakterystyczne nazywane są również punktami, kluczowych bądź kontrolnymi CP (ang. Control Point). Cechy takie rozumiane są jako specyficzne konfiguracje pikseli układające się w struktury. Przykładami takich konfiguracji mogą być: zakończenia linii, krawędzie lub kąty. Wykrywanie struktur zazwyczaj następuje za pomocą filtracji obrazu. Fragmenty obrazu, które nie zmieniają się podczas przekształceń obrazu nazywane są punktami kluczowymi. Ważnym elementem punktów kluczowych jest to ich skalo-niezmienniczość.

#### Detektor Harris’a

Jako punkty charakterystyczne można wykorzystać narożniki wykryte na obrazach. Detektor Harrisa, który jest ulepszoną wersją detektora Moraveca z powodzeniem wykrywa narożniki. Koncepcja detektora Moraveca polega na przeszukiwaniu obrazu z wykorzystaniem okna przeszukiwania ROI (ang. Region of Interest). Gdy podczas przeszukiwania w jednym kierunku zostanie zauważona duża zmiana w jasności pikseli obszar zostaje zakwalifikowany jako krawędź. Jeżeli zmiana zostanie również zauważona w kierunku prostopadłym to obszar kwalifikuje się jako narożnik. Zmiany intensywności sprawdza się co 45˚ co jest istotną wadą. W detektorze Moraveca wadami są również zaszumiona odpowiedź z uwagi na binarną funkcję okna oraz minimum jako kryterium. Chcąc poprawić algorytm Chris Harris w 1988 roku zaproponował ulepszoną wersję algorytmu z powodzeniem stosowaną do dnia dzisiejszego. W detektorze Harrisa wykorzystywana jest macierz autokorelacji w postaci:

Równanie Macierz autokorelacji detektora Harrisa

gdzie I jest funkcją intensywności, a σ jest odchyleniem standardowym funkcji Gaussa. Jeżeli różnica między wyznacznikiem macierzy M, a kwadratem jej śladu przemnożonym przez stałą k ( ) będzie mniejsza od zera to punkt jest uważany za krawędź. Jeżeli zaś to punkt uważany jest za narożnik. Jeżeli punkt jest bliski zera to uważa się, że obszar nie posiada znaczących zmian. Dodatkowo dla większej dokładności algorytmu wykrywania narożników wprowadza się próg . Zaleca się również wybrania punktów z lokalnym maksimum. Detektor Harrisa wykrywa narożniki z wiele większą dokładnością niż detektor Moraveca oraz sprawdza narożniki pod każdym kątem. Algorytm może z powodzeniem zostać wykorzystywany do tworzenia panoram.

#### GFTT

Algorytm GFTT (ang. - Good Features To Track) jest kolejnym ulepszeniem jakie wprowadzili Jianbo Shi oraz Carlo Tomasi w 1994 roku dla detektora Harrisa. Na potrzeby algorytmy wymaga się by obraz wejściowy był czarno biały. Zmianie w odniesieniu do oryginalnego algorytmu polegała na zastosowaniu innego wzoru wyliczającego funkcje oceniającą. Zamiast jak w oryginale w przypadku GFTT funkcja ta ma postać

Równanie Funkcja oceniająca dla algorytmu GFTT

Dzięki swojej modyfikacji algorytm znajduje najbardziej użyteczne wierzchołki obrazu. Każdy wykryty wierzchołek sprawdzana się czy przekroczył pewien próg zmiany jasności i jeżeli tego nie zrobił zostaje automatycznie odrzucony. Dodatkowo określa się minimalny dystans pomiędzy wierzchołkami. Kryteria pozwalają wyłonić najlepiej procentujące wierzchołki, czyli intersujące punkty kluczowe.

#### FAST

Detektor FAST (ang. Features from Accelerated Segment Test) w swoim kryterium porównuje jasność piksela z jasnością pikseli oddalonych o ustalony promień. W przypadku gdy jasność większości pikseli będzie się różnić od piksela centralnego o określoną wartość to punkt kwalifikowany jest jako narożnik. Przykładowo dla promienia równego 3 piksele, jeżeli 12 z 16 pikseli będą jaśniejsze niż piksel centralny pomniejszony o pewną wartość to algorytm wykrył narożnik. Na potrzeby algorytmu promień powinien być dostatecznie mały (od 2 do 6 pikseli) oraz wartość progu kwalifikacji i ilości pikseli odpowiednio przeskalowana do promienia. Algorytm FAST osiąga przeciętne wyniki w wykrywaniu punktów kluczowych, lecz jego zaletą jest szybkość.

#### SIFT

Algorytmem, który wykorzystuje punkty charakterystyczne na którego należało by zwrócić szczególną uwagę jest zaproponowany przez David Lowe algorytm SIFT (ang. Scale Invariant Feature Transform) [8]. Jest to jeden z bardziej popularnych algorytmów, dlatego jego działanie zostało przybliżone w tym akapicie. Algorytm SIFT realizowany jest w czterech krokach. Kroki zostały wpierw przedstawione, a następnie szczegółowo opisane. Pierwszy krok nazwany „scale space extrema detection” wykrywa punkty ekstremalne na dwóch obrazach. W kroku drugim „accurate keypoint location” następuje dokładna lokalizacja punktów charakterystycznych. Następnie przypisuje się orientację wykrytym punktom w przestrzeni. Krok nazwany jest „keypoint orientation assignment”. Czwarty etap zawiera tworzenie deskryptorów dla punktów charakterystycznych. Szczegółowe omawianie algorytmu należy zacząć od procesu skalowania obrazu wejściowego do różnych wielkości. Dla każdego obrazu w skali zostaje użyty filtr Laplace’a, który pozwala uzyskać kontury obiektów , obliczany za pomocą odjęcia dwóch obrazów rozmytych filtrem Gaussa z różnymi parametrem σ. Zazwyczaj większe rozmycie obrazu tworzy się poprzez zwiększenie potęgi, do której zostanie podniesiona stała k. Dla ułatwienia zrozumienia poniżej opisano wzór

Zazwyczaj w za parametr odchylenia standardowego przyjmuje się . Stała podnoszona do kolejnych potęg. Następnie przez binaryzację wykrywa się lokalne maksima i minima. Taka operacja powoduje wykrycie ogromnej ilości punktów ekstremalnych, dlatego należy zastosować dwa kryterium, które odfiltrują najlepsze możliwe punkty charakterystyczne. Pierwsze kryterium odrzuca płytkie minima bądź niewielkie maksima za pomocą rozwinięcia funkcji w szereg Taylora. Dzięki temu odrzucone punkty, które powstały na obszarze np. nieba. Drugie kryterium sprawdza czy punkt nie leży na odcinku za pomocą detektora Harris’a. Po tym etapie wyłonione zostają punkty niezmiennicze. Algorytm przypisuje punktom orientację co powoduje niezmienniczość również względem orientacji. Można więc ustawić obrazy tak by odpowiadające sobie punkty charakterystyczne na każdym obrazie wskazywały jednakowy kierunek. Realizowane jest to poprzez wyznaczenie gradientu w punkcie charakterystycznym. Następnie wyznacza się gradienty w małym otoczeniu punktu charakterystycznego i z odpowiednio wyważonych gradientów tworzy się histogram. Z histogramu powstaje deskryptor, który pozwala wyznaczyć orientację obszaru przetwarzanego w końcowym etapie. Etap końcowy polega na tworzeniu ostatecznych deskryptorów. Najczęściej przetwarzaniu podlega rozmyty wejściowy obraz. Etap przypomina ten poprzedni, lecz otoczenie punktu charakterystycznego jest dużo większe i podzielone na cztery obszary. W każdym obszarze wyznacza się osobne deskryptory co kończy algorytm. Uzyskuje się w ten sposób punkty charakterystyczne wraz z dokładną orientacją.

### Podsumowanie tworzenia panoramy na potrzeby systemu TSR

Chcąc zakończyć tematykę tworzenia panoramy należało by wspomnieć w jaki sposób za pomocą metod klasycznych oraz tych bazujących na punktach kluczowych zostaje utworzony jednolity obraz panoramiczny. Nietrudno zauważyć, że w każdej z metod wykryte cechy pozwalają dopasować obrazy do siebie. Następnie nakłada się jeden obraz na drugi w taki sposób by występowała największa liczba pasujących elementów np. wykrytych linii, krawędzi, narożników czy deskryptorów. Obrazy zostają scalone w jeden tzn. dokonuje się ich fuzji. Zazwyczaj scalenie polega na przysłonięciu części obrazu drugim obrazem. Istnieją również prace, które starają się wspólny obszar zapełnić informacjami z obu obrazów, lecz takie rozwiązanie może spowodować pogorszenie się jakości obrazu. Tworzenie panoramy na potrzebę systemu wykrywania znaków zostało zaimplementowane i przedstawione w akapicie Program.

## Segmentacja obrazów

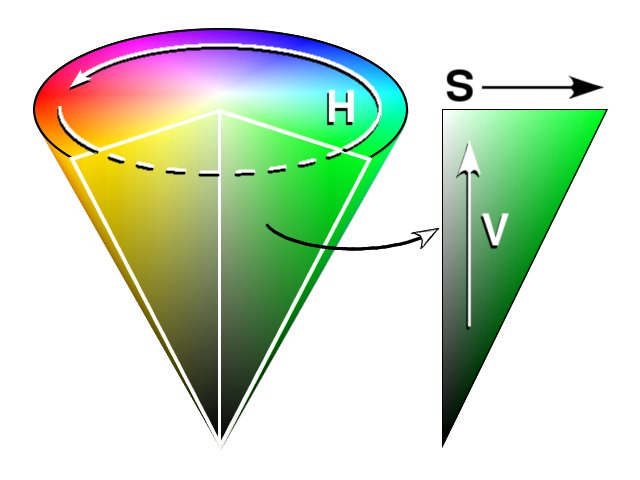
Niezależnie od tego w jaki sposób będzie rejestrowany obraz pierwszym etapem rozpoznania na nim znaków będzie znalezienie ich położenia na obrazie. Segmentacja obrazów polega na znalezieniu obiektów na obrazie i podzieleniu go według nich. Zazwyczaj na obrazie wyszukuje się tylko obiekty najistotniejsze ze względu na badany problem. W przypadku systemu TSR segmentacja obrazów polega na znalezieniu na obrazie znaku drogowego. Wykrywanie obiektów na obrazach może się odbywać na wiele sposobów. Kilka metod wykrywania i rozpoznawania obiektu zostały przedstawione poniżej. Wpierw zostało omówione Wykrywanie na podstawie koloru. W kolejnym akapicie omówiono Wykrywanie na podstawie kształtu. Następnie przedstawiono Wykrywanie na podstawie tekstury. Jedne z bardziej skutecznych jest Podejście hybrydowe próbujące łączyć ze sobą podejścia na podstawie koloru, kształtu i tekstury. Na sam koniec zostały przytoczone Inne podejścia stosowane do problemu rozpoznawania znaków drogowych.

### Wykrywanie znaków na podstawie koloru

Wykrywanie obiektów na podstawie koloru najczęściej stosuje się w przypadku kolor obiektu jest nieskomplikowany i dobrze określony. By wykrywanie na podstawie koloru mogło się odbyć, obiekt, który chcemy wykryć powinien posiadać z góry określone kolory. Znaki drogowe posiadają taką własność oraz projektowane są w taki sposób by ich kolor znacznie odróżniał się od tła. Najpopularniejszą metodą wykrywania obiektów na podstawie koloru jest progowanie obrazu w określonej barwie. Zaletą progowania jest jego szybkość. Ważnym jest by kolor znacząco odróżniał się od tła. Znaczna część przedstawionych prac np. [9] [10] wykorzystuje progowanie obrazu w kolorze czerwonym by wstępnie wyszukać regiony, gdzie mogą znajdować się znaki. Określenie koloru znaku może być zrealizowane za pomocą kilku sposobów. Zazwyczaj wykonuje się to ręcznie przez pobranie próbek z kilkunastu wzorcowych obiektów. Następnie wyznaczony zostaje średni kolor i możliwy zakres odchylenia od wzorca. Warto zauważyć, że progowanie jest mało odporne na wszelkiego typu zakłócenia, ponieważ kolor obiektu w dużej mierze zależy od oświetlenia. By zaradzić problemowi zmiany oświetlenia w zależności od warunków część algorytmów wykorzystuje możliwość zmiany przestrzeni barw.

#### Zmiana przestrzeni barw

Przestrzeń barw RGB (ang. Red, Green, Blue) jest bardzo zmienna w zależności od warunków panujących na drodze. Pogoda, pora dnia lub mnogość różnokolorowych źródeł światła mogą spowodować zmianę odbieranych przez kamerę kolorów. Część algorytmów przed progowaniem wstępnie przetwarza obraz stosując techniki tj. rozciągnięcie histogramu. Metody te są jednak mało skuteczne, dlatego część naukowców badających temat TSR postanowiło przejść z przestrzeni barw RGB na HSI (ang. Hue, Saturation, Intensity) lub HSV (ang. Hue Saturation Value) [11], [12], [13]. W przestrzeni barw HSI każda z trzech wartości odpowiada za inną składową koloru. Hue opisywany jest za pomocą koła, gdzie jego wycinek odpowiada za dany kolor. Saturation wyznacza nasycenie barwy. Czym wartość saturacji jest mniejsza tym kolor jest bledszy. Ostatnia wartość Value odpowiedzialna jest za jasność koloru. Rysunek 4 graficznie przedstawia trzy składowe przestrzeni HSI.



Rysunek Składowe przestrzeni barw HSV

Modele barw HSI i HSV pozwalają w pewnym stopniu zniwelować wahania spowodowane zaburzeniami oświetleniowymi. Niestety nie są one wolne od wad. Nie uwzględniają temperatury barwy. Przestrzeń barw LCH (ang. Lightness, Chroma, Hue) uzyskana za pomocą modelu CIECAM97 pozwala uwzględnić temperaturę barw. Warto wspomnieć również o przestrzeni YUV, która posiada trzy kanały. Jeden kanał Y dostarczający luminację, czyli jasność obrazu i dwa kanały U i V kodującą chrominancję, czyli barwę. Ta przestrzeń zdobyła największą popularność, gdy wprowadzano telewizory kolorowe. Telewizory czarno-białe odbierały jeden kanał, którym przesyłano wartość chrominancji, a nowsze odbiorniki odbierały wszystkie trzy składowe mogąc tym samym odtworzyć obraz kolorowy. Algorytm z pracy [14] bazujący na splotowej sieci neuronowej zaprzęgniętej do problemu lokalizacji znaków wykorzystuje właśnie przejście na taką przestrzeń barw. By opis był pełny należy zaznaczyć, że część autorów w swoich pracach stwierdza, że zmiana przestrzeni barw jest zbędna, ponieważ unormowana przestrzeń RGB w przeciętnych warunkach jest wystarczająca, a koszt obliczeniowy związany ze zmianą przestrzeni barw jest zbyt duży.

|  |
| --- |
| Obraz zawierający zewnętrzne, droga, niebo, samochód  Opis wygenerowany automatycznie |

Rysunek Progowanie w przestrzeni barw HSV dla koloru czerwonego

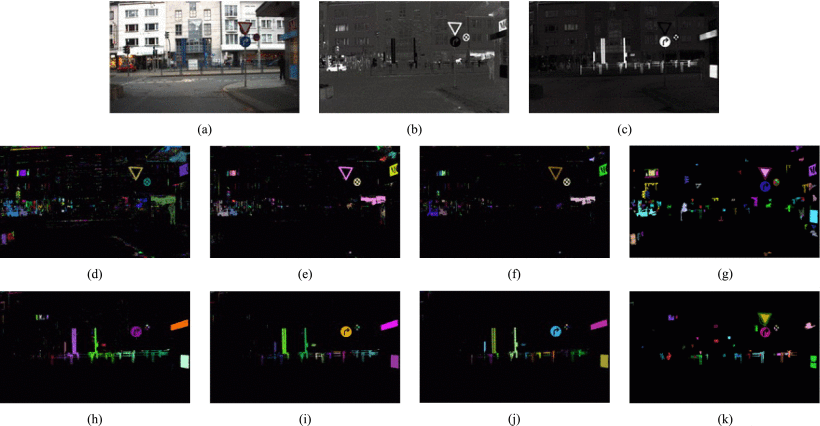
Rysunek 5 przedstawia zdjęcie poddane progowaniu w kolorze czerwonym. Obraz oraz parametry progowania zostały dobrane doświadczalnie. Progowanie uwidoczniło znak drogowy, lecz również inne elementy. Na takim obrazie albo wyszukuje się figur, które najbardziej są zbliżone do znaków drogowych np. za pomocą metody nazwanej Transformata Hougha opisanej w znajdującym się nieco niżej akapicie lub można starać się odfiltrować elementy niewłaściwe. Drugie rozwiązanie wprowadzili autorzy pracy [15].

#### AdaBoost

AdaBoost to oparty na boostingu algorytm zaprezentowany 1996 roku przez Yoav Freund i Robert Schapire [15]. Główna idea polega na stworzeniu z wielu słabych klasyfikatorów jednego silnego klasyfikator. Ciekawe rozwiązanie przedstawiono w pracy [10], w której wyszukanie znaków opiera się na kolorach, ale nie na samym progowaniu. Algorytm używa zagnieżdżonej kaskady klasyfikatora Real AdaBoost, którą szkoli się za pomocą cech LRP (ang. Local Rank Pattern). Do obliczenia cech LRP wykorzystuje się siedem typów obrazów, gdzie każdy obraz tworzy się przez odjęcie jednego typu koloru (np. czerwonego, niebieskiego zielonego itp.) W ogólności algorytm AdaBoost jest bardzo popularny i został wykorzystany w większej liczbie prac.

#### Model prawdopodobieństwa barwy

Algorytm korzystający z modelu prawdopodobieństwa barw do wykrywania znaków drogowych został przedstawiony w pracy [16]. Algorytm tworzy mapę prawdopodobieństwa występowania znaku drogowego na podstawie poszukiwania kolorów znaków drogowych w obrazie. Następnie w regionach, gdzie jest najwięcej poszukiwanych kolorów algorytm stara się wykryć znak. Dodatkowo poszukiwanie koloru następuje w przestrzeni kolorów Otha, ponieważ taka przestrzeń lepiej sprawdziła się w eksperymentach prowadzonych dla pracy. Skanowanie w przestrzeni Otha zostało przedstawione na Rysunek 7.



Rysunek 6 Tworzenie mapy prawdopodobieństwa w przestrzeni Otha

Na wyżej pokazanym Rysunek 4 obraz (a) jest wejściowym obrazem w kolorze. Obrazy Rysunek 4(b) i (c) są mapami prawdopodobieństwa uzyskanymi odpowiednio dla kolorów czerwonego i niebieskiego. Jak można zauważyć, czerwone i niebieskie piksele na oryginalnym występują w dużej ilości i mają spore natężenie. Zastosowanie map prawdopodobieństwa ułatwia wykrywanie znaków przez zwiększenie kontrastu między znakami drogowymi, a tłem. Aby algorytm pracował w czasie rzeczywistym, obliczenia zostały przyspieszone za pomocą wprowadzenia tabeli wyszukiwania LUT (ang. Look Up Table). Podczas wykrywania obliczony zostaje indeks każdego piksela według wartość RGB, a następnie znajduje odpowiadające mu prawdopodobieństwo w tabeli LUT. Autorzy algorytmu twierdzą, że po wprowadzeniu tabeli czas obliczenia map prawdopodobieństwa dla obrazu w rozdzielczości 1360×800 może został zredukowany z kilku minut do około 30 ms komputerze czterordzeniowym procesorem Intel 3,4 GHz, 4 G RAM. Dodatkowo autorzy wykorzystali detektor regionów MSER (ang. Maximally Stable Extremal Regions) aby na mapach prawdopodobieństwa zlokalizować miejsca występowania znaków. Działanie detektora MSER przedstawione zostało na obrazach od (d) do (f) dla progowania w kolorze czerwonym i od (h) do (j), dla koloru niebieskiego. Obrazy (g) i (k) sumę obrazów dostarczonych przez detektor. Na obrazach (g) i (k) następuje wykrycie obszarów ze znakami przez ztworzenie map prawdopodobieństwa. Po wykryciu regionów ze znakami następuje klasyfikacja regionów do poszczególnych podklas poprzez maszynę wektorów nośnych przeszkolonych dla cech HOG z uwzględnieniem koloru. Algorytm HOG został przedstawiony w podrozdziale opisującym wykrywanie znaków na podstawnie tekstury, dlatego w tym miejscu nie został opisany.

### Wykrywanie na podstawie kształtu

Powszechnym podejściem w wykrywaniu obiektów w tym znaków drogowych na obrazach jest wykrywanie poprzez kształt. Podobnie jak w przypadku koloru tak i w przypadku wykrywania na podstawie kształtu ważnym atutem obiektu, który należy zlokalizować musi być konkretnie określony kształt. Jak dla znaków drogowych taki kryterium jest spełnione. Znaki informacyjne są kwadratowe, nakazu lub zakazu okrągłe, a ostrzegawcze trójkątne. Wykrywanie na podstawie kształtu nie jest również wolne od wad. Percepcja kształtów różni się od kąta obserwowania. Innym problemem jest fakt, iż podczas obserwacji znaki mogą być częściowo przysłonięte lub na ich część może padać cień, co zmienia ich wygląd dla większości algorytmów. Ważnym jest by zastosowany detektor radził sobie z takimi problemami. Wykrywanie figur geometrycznych na obrazie poprzez kształt może być zrealizowane na kilka sposobów. Te ważniejsze dla pracy zostały przedstawione poniżej.

#### Wykrywanie krawędzi

Wykrywanie krawędzi jest popularną metodą uwidaczniania obiektów na obrazie. W tym miejscu należy objaśnić, że krawędź identyfikowana jest jako przejście z obszaru ciemniejszego do jaśniejszego bądź na odwrót, jest to granica pomiędzy dwoma obszarami o różnych jasnościach. Takie podejście wymaga ustalenia konkretnego progu zmiany jasności powodującego wykrycie. Część metod bazuje na operatorach gradientowych ustalając próg lokalny, zamiast globalnego. Takie metody wykorzystują zmiany pierwszej lub drugiej pochodnej obrazu w skali szarości. Powszechnym rozwiązaniem wykrywania krawędzi jest Detektor Canny, który został omówiony poniżej. Istnieją również inne metody. Przykładami takich metod jest Krzyż Robertsa oraz Operatory Sobela. Oba algorytmy sprawdzają różnicę między sąsiednimi pikselami w obrazie przez splot pewnej macierzy z obrazem. Algorytmy bazujące na takich operatorach przeznaczone do problemu TSR przedstawione zostały np. w pracy [17].

#### Detektor Canny

Jak wcześniej zostało wspomniane jedną z popularniejszych metod wykrywania krawędzi jest metoda Canny zaprezentowana w Johna F. Canny w 1986. Metoda przedstawia się następująco. Wpierw następuje filtracja obrazu za pomocą filtru Gaussa. Powoduje to wstępnie rozmycie obrazu i odfiltrowanie małych zakłóceń. Następnie za pomocą np. operatora Sobela wykryty zostaje gradient zmian poziomych i pionowych każdego punktu na obrazie. W kolejnym kroku wyznacza się na podstawie wykrytych gradientów dwa parametry:

|  |
| --- |
| . |

Rysunek Parametry krawędzi dla algorytmu Canny

Wartośćoznacza długość, a kąt detekcji krawędzi. Niekiedy kąt zaokrągla się do wartości liczonych co 45˚. Piksele, które nie łączą się z żadną pobliską krawędzią zostają odrzucone. Piksele umieszone blisko zakończeń wykrytych krawędzi zostają połączone w jedną krawędź.

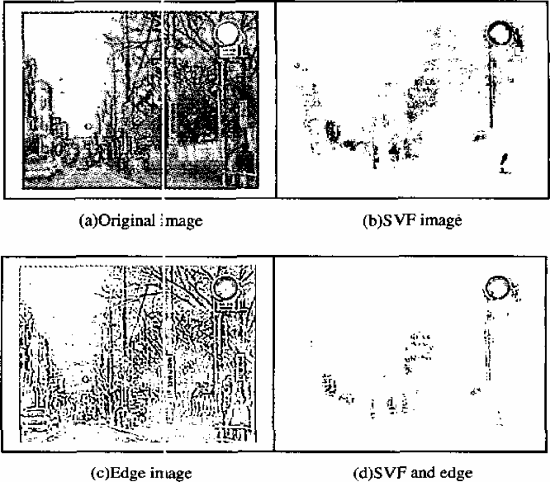
|  |
| --- |
| Obraz zawierający zewnętrzne, droga, niebo, samochód  Opis wygenerowany automatycznieObraz zawierający przyroda, deszcz  Opis wygenerowany automatycznie |

Rysunek Detekcja krawędzi poprzez algorytm Canny

Rysunek 8 przedstawia obraz uzyskany za pomocą detekcji krawędzi Canny. Na tak przetworzonym obrazie należy odszukać obiekty, które nas interesują. Dla problemu TSR są to: trójkąty, kwadraty i okręgi. Detektor Canny dzięki swoim właściwością doskonale nadaje się do uwidaczniania kształtów, dlatego jest to algorytm, po który najczęściej się sięga, gdy wymagane jest wykrycie konturów.

#### SVF

SVF (ang. Simple Vector Filter) jest algorytmem pozwalającym w pewnym stopniu odfiltrować z obrazu, na którym wykryto krawędzie nieistotne elementy. Metoda została zaproponowana w artykule [12]. W podanej pracy wpierw następuje wyodrębnienie kształtów przez zastosowanie filtra krawędziowego bazującego na zmianie jasności. Niestety takie działanie wyodrębnia dużo niepożądanych elementów sceny. Należało więc z wszystkich znalezionych elementów zostawić te najbardziej podobne do znaków, a resztę usunąć. Największym problemem są obiekty, których kształt zbliżony jest do kształtu znaków. By poradzić sobie z tą przeszkodą autorzy zaproponowali autorski filtr SVF uwzględniający również kolor znalezionego obiektu. Zaletą proponowanego filtru jest jego szybkość. Dokładny opis przebiegu algorytmu jest bezcelowy, ponieważ można go doczytać w podanej pracy. Skuteczność metody przedstawiono na Rysunek 5. Na obrazie (a) pokazano obraz oryginalny przekonwertowany do skali szarości i zbinaryzowany. Obraz (b) przedstawia działanie filtra SVF nałożonym na obraz (a). na obrazie (c) uwidoczniono krawędzie z obrazu (a). Ostatni obraz (d) przedstawia zastosowanie filtra SVF na obrazie (c). Jak można zauważyć metoda dobrze sprawdza się do filtracji i może poprawić wykrywanie znaków przy wykrywaniu na podstawie kształtu.



Rysunek Wykrywanie znaków z zastosowaniem filtra SVF

#### Wykrywanie rogów znaków jako punków charakterystycznych

Wykrywanie rogów może być przydatne zwłaszcza dla znaków ostrzegawczych (trójkąty) i informacyjnych (kwadraty). Wykrywanie narożników może odbywać się np. za pomocą detektora Harrisa opisanego w wcześniejszym akapicie zatytułowanym Detektor Harris’a. Jeżeli nastąpi wykrycie rogów w konkretnej konfiguracji adekwatnej dla znaku można domniemywać, że w tym miejscu wystąpił znak i przystąpić do klasyfikacji znaku.

#### Transformata Hougha

Algorytmem, na który należało by również zwrócić szczególną uwagę jest transformata Hougha. Pozwala ona odnajdywać regularne kształty na obrazie. Została zaprezentowana w 1962 roku jako metoda wykrywania linii prostych, lecz udało się go zastosować dla wykrywania innych kształtów takich jak okręgi, kwadraty czy trójkąty [18]. W 1993 roku Anagnou przestawił udoskonaloną metodę pozwalającą uzyskanie większej rozdzielczości kątowej oraz radialnej znajdowanych linii. Omówienie metody zostało zrealizowane na przykładzie wykrywania kształtu kwadratowego. W tej metodzie każda prosta opisana jest równaniem . Początkowo zostaje zaimplementowana tablica (komórki nazywa się akumulatorami) z wartościami zerowymi. Następnie każdy punkt obrazu przekształcany jest w dyskretną krzywą sinusoidalną w przestrzeni Oρϴ. Oblicza się wartości parametru dla kątów w zakresie ϴ = (0, 360˚). Obliczenie następuje przez dodanie wartości jeden do liczby oznaczającej wartość odpowiedniego akumulatora. Po wykonaniu inkrementacji dla wszystkich punktów obrazu, wybierane zostają akumulatory z największą wartością . Położenie komórki określają położenie i kąt odcinka zaś wielkość jego długość. Według podobnej zasady można również wykryć okręgi na obrazie. Transformacja Hougha odnosi sukcesy w wykrywaniu linii czy okręgów, lecz do wykrycia innych kształtów jak np. trójkąty wymaga skomplikowanych obliczeń i jest obciążająca pamięciowo. Metody transformaty Hougha użyto np. w pracy [19] do rozróżnienia typu znaku. Praca ta została bliżej przedstawiona w rozdziale klasyfikacji znaków i metodzie SVM.

### Wykrywanie na podstawie tekstury

Podczas wykrywania obiektów na podstawie koloru i kształtu uzyskanie dobrych wyników może być problematyczne, dlatego zaczęto szukać innych rozwiązań mogących poprawić skuteczność wykrywania. Rozwiązaniem mające duży potencjał było wykrywanie na podstawie tekstury. Omawiając je należały wyjaśnić, że tekstura to graficzny odpowiednik faktury. Faktura obiektów w rzeczywistości bardzo wpływa na postrzeganie ich przez człowieka, dlatego część grup zajmujących się odnajdywaniem obiektów na obrazie podejmowało próby skonstruowania algorytmu bazujących na teksturze. Kilka takich metod zostało przedstawione poniżej.

#### Histogram gradientów zorientowanych

HOG (ang. Histograms of Oriented Gradients) - histogram gradientów zorientowanych jest deskryptorem obrazu pozwalającym na znalezienie kształtu i wyłonienie obiektu [20]. Idea algorytmu polega na podzieleniu obrazu przekonwertowanego do skali szarości na małe fragmenty (np.8x8 pikseli) i obliczeniu dla każdego piksela w takim fragmencie różnicy jasności (gradientu) pomiędzy pikselami sąsiadującymi. Następnie dla każdego z fragmentu wyznacza się histogram gradientu. W przeciwieństwie do wcześniej przedstawionego algorytmu SIFT, HOG oblicza deskryptory w równomiernie oddalonych fragmentach obrazu. Zastosowanie lokalnej normalizacji kontrastu w nakładających się na siebie regionach poprawia skuteczność wykrywania obiektów [21]. Zaproponowany w pracy [22] wariant HOG (HOGv) posiada dwie modyfikacje poprawiające wykrywanie. Pierwszą z nich jest uwzględnienie zarówno wrażliwych jak i niewrażliwych na kontrast orientacji gradientów, w taki sposób, że bardziej szczegółowe lokalne informacje o znakach mogą być włączone do zgromadzonych histogramów. Drugą modyfikacją jest to, że każdą komórkę normalizuje się odpowiednio z czterema sąsiednimi blokami. Znormalizowane histogramy komórki są wymiarowo zmniejszane za pomocą strategii podstawowej analizy komponentów (PCA). Takie działanie ma na celu usunięcie nadmiarowych informacji.

#### Lokalne wzorce binarne

LPB (ang. local binary patterns) zostało opisane w 1994 roku jako algorytm wykrywania obiektów na obrazie [23]. Algorytm odniósł duży sukces w wykrywaniu twarzy, dlatego starano się go zaimplementować do wykrywania innych obiektów. Jego idea jest podobna do algorytmu HOG, ponieważ na początku obraz zostaje podzielony na mniejsze obszary. Następnie porównuje się, każdy piksel z wszystkimi pikselami sąsiadującymi w ustalonej kolejności. W zależności czy wartość piksela porównywanego jest większa lub równa czy mniejsza od piksela centralnego w osobnej tablicy zostaje wpisana odpowiednio wartość 1 lub 0. Takie rozwiązanie niesie ze sobą zaletę, że gdy zostaje zmieniona globalna jasność obrazu, wartości tablic [0,1, (…)] się nie zmienią i skuteczność algorytmu nie spadnie. Następnie podobnie jak w algorytmie HOG w fragmentach zostaje utworzony histogram, a następnie zostaje ustalony gradient całego okna. Po wyznaczeniu gradientów w każdym oknie następuje połączenie podobnych gradientów mogących tworzyć jeden obiekt. Jeżeli odnalezione pole wyznaczone przez gradienty jest podobne do jednego z obiektów przechowywanych w bazie algorytm kończy działanie. LBP dla wykrywania znaków drogowych zaproponowano w pracy [19]. Rozwiązanie w zbogacone jest o ustalenie na początku działania potencjalnych obszarów poszukiwań znaków za pomocą okna przesuwanego o niewielkich rozmiarach. Etap nazywa się filtrowaniem zgrubnym i używa algorytmu LDA. Następnie podczas dokładnego filtrowania weryfikuje się okna odnalezione w etapie poprzednim za pomocą metody NMS (ang.  Non-Maximal Suppression) wykonywane jest filtrowanie okien mogących odnosić się do tego samego obszaru. Po odnalezieniu obszarów, gdzie możliwe jest wystąpienie znaków drogowych ich klasyfikacja wykonywana jest za pomocą maszyny wektorów nośnych. Maszyna ta została opisana w następnym rozdziale.

### Podejście hybrydowe oraz inne podejścia

Podejście hybrydowe w ogólności polega na połączeniu metod opartych na kolorze, kształcie i fakturze. W większości prac na początkowym etapie następuje progowanie w określonych kolorach co zawęża obszar poszukiwań, a następnie wykorzystuje się podejścia oparte na kształcie i fakturze w celu poprawy wydajności wykrywania.

#### Detekcja krawędzi wzbogacona o Prosty Filtr Wektorowy

Przykładem metody hybrydowej stworzonej do segmentacji obiektów na obrazie jest metoda bazująca na wykryciu krawędzi oraz późniejszym zastosowaniu filtru z SVF (ang. Simple Vector Filter). Metoda ta została zaproponowana przez autorów artykułu [12]. W podanej pracy wpierw obraz kolorowy dostarczony przez kamerę konwertowano się do skalali szarości. Następnie starano się wyodrębnić kształty obiektów poprzez progowanie obrazu albo detekcję krawędzi poprzez metodę bazującą na zmianie jasności. Niestety, takie działanie wyodrębnia wszystkie obiekty sceny, dlatego należało odróżnić znaki drogowe od tła. Największym problemem były obiekty, których kształt zbliżony był do kształtu znaków. By poradzić sobie z tą przeszkodą autorzy zaproponowali autorski filtr SVF uwzględniający kolor znalezionego obiektu. Zaletą proponowanego filtru jest jego szybkość. Dokładny opis przebiegu algorytmu można go doczytać w podanej pracy. Skuteczność SVF w odfiltrowaniu znaków drogowych od reszty obiektów przedstawiono na Rysunek 9.

|  |  |
| --- | --- |
| Obraz zawierający zewnętrzne, drzewo  Opis wygenerowany automatycznie   1. - Zbinaryzowany | Obraz zawierający drzewo, niebo, zewnętrzne  Opis wygenerowany automatycznie   1. - (a) z SVF |
| (c) – po wykryciu krawędzi | 1. - (c) z SVF |

Rysunek Wykrywanie z zastosowaniem filtra SVF

Obraz (a) przedstawia obraz, który został zbinaryzowany, (b) przedstawia obraz (a) po nałożeniu filtru SVF. (c) oryginalny z wykrytymi krawędziami, a obraz (d) obraz analogicznie jak wyżej obraz (c) po nałożeniu filtra SVF. By sprawdzić czy w wykrytym regionie istnieje znak zastosowano algorytm genetyczny opisany w akapicie jako jeden z przykładów w akapicie Algorytmy Genetyczne.

#### Algorytmy Genetyczne na potrzeby segmentacji

Algorytmy genetyczne są popularnym podejściem w segmentacji i klasyfikacji znaków drogowych. Działanie GA (ang. Genetic Algorithm) jest iteracyjne iteracyjnie. W każdej iteracji stara się znaleźć lepsze rozwiązanie poprzez krzyżowanie i mutacje znalezionych rozwiązań. Początkowe wartości rozwiązań są losowane ze zbioru możliwych rozwiązań. Algorytm kończy działanie, gdy wyszukane rozwiązanie jest wystarczająco skuteczne.

Rozwiązanie problemu klasyfikacji znaków drogowych z pomocą algorytmu genetycznego przedstawione zostało w wcześniej opisywanej pracy [13] Algorytm skupia się na wgrywaniu znaków okrągłych na obrazach dostarczonych przez filtr krawędziowy i SVF. Oczywiście znaki okrągłe można wyszukać na obrazach przez znalezienie okręgów. W TSR dużym problemem jest dynamicznie zmieniający się promień okręgu i właśnie do znalezienia najlepszego promienia użyto algorytmów genetycznych. W metodzie prezentowanej w artykule każdy osobnik złożony jest z wektora cech. Cechy są liczbami rzeczywistymi o odpowiadają współrzędnym wyszukiwania i promieniowi znaku. Dodatkowo algorytm zawsze replikuje najlepsze rozwiązanie. W artykule również proponują modyfikacje algorytmu genetycznego nazwaną Step-GA starającą się testować kolejne pokolenia na obrazach wybranych sekwencji filmowej z ustaloną częstotliwością. By poprawić zdolności wykrywania znaków przez algorytm autorzy zaproponowali prostą metodę śledzenia znaku. W większości przypadków podczas jazdy samochodem znak drogowy wykryty na środku ekranu będzie się powiększał i przesuwał w okolice prawego górnego rogu. Jeżeli algorytm wykryje obszar, który będzie zachowywał się w taki sposób dopiero wtedy może być on kwalifikowany jako znak drogowy. Takie działanie pozwala odfiltrować dużą część regionów, w których wystąpiły błędy drugiego rodzaju. Jednak mimo dużej ilości zalet metody w trakcie jej publikacji, czyli roku 2002 była dość niedoskonała. Autorzy nie testowali swojego algorytmu na żadnym międzynarodowym benchmarku, dlatego ciężko ocenić skuteczność przedstawionej metody.

Algorytm genetyczny do znajdowania znaków został zaproponowany w pracy [24]. Zaproponowana metoda działa w następujący sposób. Wpierw obraz konwertowany jest z przestrzeni RGB do przestrzeni HSV. Następnie następuje progowanie w kolorze czerwonym. Autorzy w przeciwieństwie do klasycznej metody progowania nie używają konkretnie określonych wartości progów W celu wybrania obszarów czerwonych zostały przygotowane tablice LUT (Look Up Table). W ten sposób błędy klasyfikacji pikseli spowodowane progowaniem powinny zostać zrekompensowane. Przygotowano dwie takie tablice, dla barwy oraz nasycenia zaprezentowane na Rysunek 10.

|  |  |
| --- | --- |
| Obraz zawierający obiekt  Opis wygenerowany automatycznie   1. - Barwa | 1. – Nasycenie |

Rysunek Tablice LUT

Po splocie każdego piksela z wartościami tablicy obraz zostaje znormalizowany do wartości 255. Dokładna klasyfikacja nie jest konieczna, ponieważ tylko kształty obiektów zamkniętych będą brane pod uwagę. Na tak przygotowanym obrazie zaczyna pracę algorytm genetyczny. Metoda GA musiała zostać zbilansowana pomiędzy przeszukiwaniem całej przestrzeni obrazu, a miejscem, w którym najbardziej prawdopodobne było wystąpienie znaku. Inaczej mówiąc ryzykiem, z którym musi mierzyć się każdy algorytm genetyczny jest zatrzymanie się w lokalnym ekstremum. Należy więc unikać przedwczesnej zbieżności. Jest to odpowiednik utraty genetycznego bogactwa gatunku. Zapis kodu genetycznego przedstawiony został za pomocą Równanie 3 Model znaku w zadanej odległości i prostopadłego do osi optycznych.

|  |
| --- |
| Gdzie:  – przemieszczenie poziome,  – przemieszczenie pionowe,  – skala pozioma,  – skala pionowa,  – rotacja pozioma. |

Równanie Model znaku

Rozważanymi modyfikacjami była zmiana pozycji oraz skali znaku spowodowana dystansem, oraz kątem w przypadku, gdy znak nie byłby prostopadły do osi. W klasycznym algorytmie genetycznym, populacja początkowa jest generowana losowo, lecz w tym przypadku możemy przyspieszyć działanie ustawiając populacje początkowe z większym prawdopodobieństwem w miejscach, gdzie zostały odnalezione uwidocznione obiekty na obrazie dostarczonym przez progowanie. Stała liczba osobników jest przyporządkowywana do każdego obiektu. W ten sposób obecność wystarczającej liczby osobników może być zagwarantowana pomimo obecności większych obiektów. Reguła dopasowania została oparta na odległości Hausdorffa. Odległość ta wskazuje jak dwa kształty różnią się od siebie. Jeśli mamy dwa zbiory punktów A = {a1, ..., am} i B = {b1, ...bn} to odległość Hausdorffa zdefiniowana jest następująco:

|  |
| --- |
|  |

Równanie Odległość Hausdorffa

Funkcja jest to bezpośrednia odległość Hausdorffa. Definiuje ona punkt należący do zbioru , który jest najdalej, względem wybranej normy, od każdego z punktów zbioru . Wariacja jest częściową odległością Hausdorffa, w której tylko odległości jest brane pod uwagę. Miara jest nieczuła na zakłócenia. W opisanym przypadku oba zbiory zawierają krawędzie obiektów wykrytych w poprzednim kroku oraz transformaty zakodowane w chromosomie każdego osobnika. Następnie obliczana jest transformata odległości oraz liczba punktów, których wartość jest mniejsza od progu. Dopasowanie jest relacją pomiędzy liczba punktów, a całkowitą liczbą punktów modelu. Kolejna zaletą tej funkcji dopasowania jest możliwość zatrzymania tworzenia pokoleń, jeśli wartość jest wystarczająco wysoka.

Obraz zawierający budynek, zdjęcie

Opis wygenerowany automatycznie

Rysunek Cztery wyselekcjonowane pokolenia GA

Rysunek 11 przedstawia cztery pokolenia algorytmu genetycznego. Obraz (a) przedstawia pierwsze pokolenie. Jak można zauważyć część osobników została zainicjalizowana w miejscu występowania znaku, lecz część z nich została rozrzucona losowo. Na czarnym tle został na fioletowo przedstawiony wzorzec znaku z obrazu, a na żółto najlepsze dotychczas znalezione rozwiązanie. Obrazy (b) i (c) są iteracjami w połowie działania algorytmu, a obraz (d) przedstawia ostatnią iterację. Można zauważyć, że algorytm dobrze dopasował obrys znaku do jego wzorca co pokazane jest na czarnym wycinku na obrazie (d). W każdym pokoleniu wybierani są rodzice, na podstawie których tworzone było kolejne pokolenie. Proces selekcji określony był poprzez wartość dopasowania rozwiązania. W tym procesie geny zawierające dobre rozwiązania były rozszerzane przez populacje. Selekcja odbywała się przy pomocy metody ruletki: prawdopodobieństwo wylosowania było proporcjonalne do znormalizowanego dopasowania wszystkich osobników. Dzięki takiemu podejściu najlepsze osobniki miały więcej potomstwa od innych. Czasami metoda ruletki miała problemy z trójkątnymi znakami. Jak wspomniano wcześniej, algorytmy genetyczne są lepsze w wyznaczaniu niż poszukiwaniu, ponieważ mogą zostać uwiezione w lokalnym ekstremum, z powodu tego, że większość osobników jest skoncentrowana w tej strefie. Trójkątne znaki często doprowadzają do tego zjawiska. Powodem jest to, że kilka osobników może nakładać dwa boki trójkąta na jeden bok modelu. Wygodnie było opóźnić zbieżność, pomimo ryzyka jej spowolnienia, aby zapewnić znalezienie globalnego ekstremum, dlatego właśnie zastosowano metodę rankingową. Jeśli powstały złe osobniki, zostały wyeliminowane w następnym doborze. Jeśli okazały się dobre, ich informacja była propagowana w populacji. Zadowalające wyniki otrzymano dla następujących parametrów:

• Poziome przemieszczenie: zakres od 0 do 384 pikseli,

• Pionowe przemieszczenie: zakres od 0 do 286 pikseli,

• Pozioma skala: zakres od 0.25 do 1.3 (od 30 do 157 pikseli),

• Pionowa skala: zakres od 0.25 do 1.3 (od 30 do 157 pikseli),

• Pozioma rotacja: zakres od -15 do 15 stopni,

• Populacja zwierała 101 osobników,

• Prawdopodobieństwo krzyżowania wynosiło 60 %,

• Prawdopodobieństwo mutacji wynosiło 3 %,

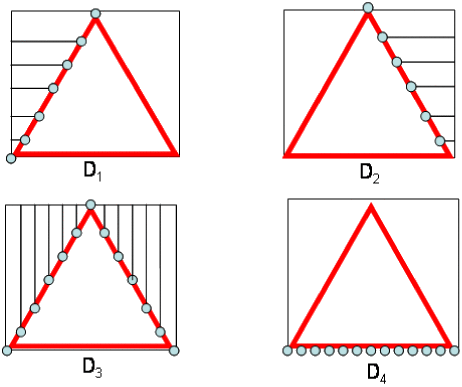
• Maksymalna liczba pokoleń wynosiła 51,

• Wartość graniczna (ang. escape value) wynosiła 80 %.

Czas pracy generowania każdego pokolenia wynosi 8.8 ms pracując na procesorze AMD Duron 1 GHz. Po znalezieniu regionów występowania znaku rozpoznanie znaku odbywa się za pomocą sieci neuronowej nauczonej przez obrazy dostarczone w przestrzeni HSI. Warto wspomnieć, że prace dotyczące algorytmów genetycznych w połączeniu z sieciami neuronowymi stosowane były od początku rozważań nad tematem TSR. Przykładem może być przytoczenie pracy [1] z 1996 roku która proponowała takie rozwiązanie.

#### Maszyna wektorów nośnych

Maszyna wektorów nośnych SVM (ang. Support Vector Machine) jest klasyfikatorem umożliwiającym określenie do jakiego zbioru należy obiekt, za pomocą jego cech. Do poprawnego działania potrzebuje zbioru danych, zawierającego poprawnie sklasyfikowane obiekty, po ich cechach. Algorytm działa na zasadzie znalezienia hiperpłaszczyzny w przestrzeni zawierającej cechy obiektów, która oddziela dwie klasy obiektów. Jeżeli taka funkcja nie istnieje w przestrzeni zadania należy wprowadzić funkcję jądra, która powiększy przestrzeń. System, który korzysta z maszyny wektorów nośnych dla problemu TSR został przedstawiony w pracy [11]. Działanie systemu polega na wykrywaniu obszary występowania znaków za pomocą progowania w przestrzeni HSI. Znaki białe wykrywane są za pomocą rozkładu achromatycznego. Każdy obszar prawdopodobnego wystąpienia znaku wpisywany jest w kwadrat. Następnie wyznacza się 20 odległości liczonych od ramki kwadratu do znalezionego koloru w środku ramki mający być obwiednią obiektu pokazanych na Rysunek 7 Odległości tworzące wektor DtB. Z tych 20 odległości tworzy się wektor DtB dostarczany jest na wejście SVM.



Rysunek Odległości tworzące wektor DtBs

Warto wspomnieć, że wyznaczony w ten sposób wektor znaków ośmiokątnych niezbyt różni się od znaków okrągłych, dlatego znaki ośmiokątne klasyfikowane są dopiero na etapie rozpoznawania znaku. Następnie przy użyciu kilku wytrenowanych maszyn SVM następuje klasyfikacja obszaru. Każda maszyna potrafi rozpoznać jeden typ obrazu. W zależności od wykrytego koloru (czerwonego, niebieskiego, żółtego lub białego) algorytm wyznacza inne maszyny do klasyfikacji. Proces rozpoznania opiera się na maszynie SVN z ziarnami Gaussa. Dużą zaletą metody jest jej odporność na rotację.

#### Histogram gradientów zorientowanych w połączeniu z maszyną wektorów nośnych

Maszyna wektorów nośnych może być również łączona z algorytmem tworzącym histogram gradientów zorientowanych. Takie podejście zaprezentowane zostało w pracy [16]. W tej pracy wykrywanie obiektów za pomocą metody HOG wspomagane jest informacją o kolorze, co jest ogromnie istotne dla wykrywania znaków drogowych. Oryginalnie algorytm HOG oblicza gradienty dla każdego kanału kolorów i przyjmuje gradient o największej normie, zaś algorytm w opisywanej pracy [16] oblicza funkcje HOG dla każdego kanału kolorów po czym łączy je ze sobą tworząc histogram. W odróżnieniu od podobnych metod, obliczenie funkcji HOG następuje na podstawie mapy prawdopodobieństwa, ponieważ mapa prawdopodobieństwa może kodować informacje o kolorze i kształcie znaku drogowego, jednocześnie tłumiąc wpływ tła. Jednak na podstawie samej mapy prawdopodobieństwa nie ma możliwości zidentyfikowania znaku. Aby rozwiązać ten problem, dodano dodatkową funkcję w postaci klasycznego algorytmu HOG. Do identyfikacji przeszkolony został wielopoziomowy klasyfikator SVM. Autorzy algorytmu przetestowali swoją pracę na zbiorze danych GTSDB, gdzie istnieją trzy kategorie znaków drogowych, dlatego wyszkolony został 4-klasowy klasyfikator SVM z dodatkową klasą tła. Za jądro klasyfikatora SVM została wybrana funkcja radialna, ponieważ najlepiej sprawdziła się w eksperymentach. Zaproponowany algorytm na bazie danych GTSDB wykrywa znaki zakazu i nakazu z 100 % poprawnością, znaki ostrzegawcze z 94.29 % poprawnością. Czas wykrycia znaków w ramce to 0,067 sekundy co jest bardzo zadowalającym wynikiem.

#### Algorytm Violi i Jonesa

Algorytm Viola–Jones zaproponowany w 2001 przez Paul Viola i Michael Jones wprawdzie został stworzony do wykrywania twarzy [24], lecz znakomicie nadaje się do wykrywania obiektów każdego typu na obrazie. Podejście dla problemu TSR zostało przedstawione w pracy [2]. W pracy jako klasyfikatora używa się algorytmu opartego na boostingu nazwanego AdaBoost (ang. Adaptive Boosting). Tworzy on z kilku mniejszych klasyfikatorów jeden silny klasyfikator. Dla ww. pracy jako słabe klasyfikatory przyjmuje się drzewa decyzyjne, w których każdy węzeł jest jedną z cech haaro-podobnych. Cechy Haara swoją nazwę zawdzięczają falką Haara zaproponowanym przez Alfréda Haara w 1909 lub 1910 roku. Cechy Haara to najprościej mówiąc pewna maska, która posiada dwa typy pikseli. Po jej nałożeniu na część obrazu sumuje się jasności pikseli należących do danego typu. Jeżeli różnica między dwoma typami jest wystarczająca cecha Haara daje pozytywną odpowiedź. Algorytm używa niewielkiej ramki przesuwanej po obrazie ROI. W każdym położeniu ramki za pomocą cech Haara stara się odnaleźć miejsce występowania znaku. Cechy Haara sprawdzane są kaskadowo co pozwala od razu odrzucić rejony, gdzie znak nie występuje. By przyspieszyć obraz zostaje scałkowany. W wcześniej już wspomnianej pracy [2] metoda pozwoliła na wykrywanie nie tylko znaków drogowych, lecz również na wykrycie samochodów a nawet rowerzystów. Ponieważ algorytm oprócz wykrywania znaków wykrywa dwa inne typy obiektów, wprowadzono dodatkowo funkcje subkategoryzacji. Wielką zaletą metody jest wykrywanie obiektów w czasie rzeczywistym. Praca, która korzysta z metody opartej na AdaBoost i cechach Haara została zaprezentowana w [25]. Dzieło pozwala na wykrywanie i śledzenie obiektów. Cechy Haara zostały również wsparte informacją o kolorach co pozwoliło na redukcję błędów pierwszej klasy z 1.6 % do 1.4 % oraz błędów drugiej klasy z 0.3 % do 0.03 %, co jest spektakularnym wynikiem. Praca [10] pozwala na śledzenie obiektów i również wykorzystuje kaskadę AdaBoost. Co ciekawe algorytm przedstawiony w pracy jest w stanie sam się doszkalać przez zbieranie pozytywnych i negatywnych próbek. Algorytm do doszkalania wykorzystuje śledzenie wsteczne. Oczywistością jest, że znaki które są bliżej są również lepiej rozróżnialne, dlatego po wykryciu znaku algorytm sprawdza poprzednie klatki filmu w poszukiwaniu wykrytego znaku w mniejszej skali i stara się samoczynnie doszkolić.

#### Symetria promieniowa

Niektóre ostatnie prace [26] i [27] wdrożyły szybki algorytm oparty na symetrii promieniowej, która dostosowany do kształtów trójkątnych, kwadratowych, diamentowych, ośmiokątnych i okrągłych. Algorytm uwzględnia gradient obrazu dostarczonego skali szarości i wykorzystuje naturę kształtów, które głosują w punkcie środkowym dla okrągłych znaków i linii głosów w przypadku regularnych wielokątów. Główną zaletą tej metody jest to, że jest ona w stanie działać w czasie rzeczywistym. Ponieważ wykrywa kształty oparte na krawędziach, algorytm jest odporny na zmiany oświetlenia.

## Klasyfikacja znaków drogowych

Klasyfikowanie znaków, wyłączając etap śledzenia, jest ostatnim etapem wykrywania znaków drogowych. Omawiana poprzednio segmentacja zawęża się obszar występowania znaku co w rezultacie prowadzi do faktu, że znak zajmuje około 70 % fragmentu odnalezionego przez segmentacje. Zadaniem klasyfikacji jest odróżnienie czy w miejscu wskazanym przez segmentacje znak rzeczywiście istnieje i jeżeli tak jest odczytanie znaku. Obecnie najpowszechniej stosowanym klasyfikatorem obrazów są sieci neuronowe. Warto jednak wspomnieć, że prócz sieci istnieją inne metody rozpoznawania typu znaku. W tej części zostały opisane metody klasyfikacji na potrzeby systemu TSR.

### Wzorcowe dopasowanie

Template matching jest klasyczną metodą bazującą na informacji wzajemnej pozwalającą odnaleźć obiekt na obrazie. Na potrzeby metody TM wzorce obiektów, które należy rozpoznać gromadzone są w bazie wzorców. Następnie zazwyczaj prostą metodą polegającą na dopasowaniu dwóch obrazów do siebie np. metodą najmniejszych kwadratów. Metoda TM jest prosta w zaimplementowaniu, lecz przez czas pracy dla kilkudziesięciu typów znaków drogowych jest nieefektywna. Szczegółowy opis metody można znaleźć w publikacji [28]

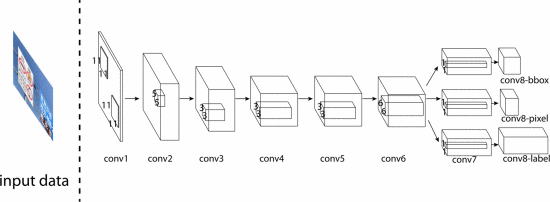
### Sztuczne sieci neuronowe

Sztuczne sieci neuronowe są zaprogramowanymi strukturami, które symulują model zdolny do wykonywania obliczeń matematycznych. Idea sztucznych sieci neuronowych inspirowana była budową ludzkich neuronów. Każda sieć zbudowana jest z warstw węzłów zwanych neuronami. Węzły połączone są ze sobą synapsami. Zazwyczaj do jednego węzła wchodzi wiele synaps. Każda synapsa posiada swoją wagę, a węzeł funkcję aktywacji. Gdy na wejściu synapsy pojawia się sygnał mnożony on jest przez wagę synapsy i podawany na węzeł. Gdy łączna wartość sygnału podanego na węzeł jest wystarczająca sygnał podawany jest dalej na kolejną warstwę. Sieci neuronowe stosowane są do wszelakich zadań, lecz na potrzeby pracy zostanie przedstawione zastosowanie w rozpoznawaniu obrazów. W takim zastosowaniu na wejściu sieci podawane są piksele obrazu, który należy rozpoznać. Zazwyczaj sieć ma określoną liczbę pikseli wejściowych, dlatego wykryty obszar przez segmentację należy przeskalować i znormalizować do tej liczby. Do normalizacji obrazu można stosować wiele metod. Najprostszą z nich jest odrzucenie co pikseli, co konkretną wartość liczbową. Rozwiązaniem, dzięki któremu traci się mniejszą ilość informacji, lecz jest bardziej kosztowny obliczeniowo jest interpolacja dwuliniowa. Już w artykule z 1996 roku [17] zostały użyte sieci neuronowe do klasyfikacji typu znaku drogowego. Istnieje ogromna liczba typów sieci neuronowych i metod usprawniających ich działanie. Sieć składająca się z więcej niż jedna warstwa ukryta nazywana jest siecią wielowarstwową. Przykładem takiej sieci jest perceptron wielowarstwowy.

#### Wielowymiarowy percreptron

MLP (ang. Multilayer perceptron) - perceptron wielowarstwowy jest najpopularniejszy rodzajem sieci neuronowych. Taka sieć składa się z jednej warstwy wejściowej i kilku warstw ukrytych. Na potrzeby klasyfikacji w węzłach sieci MLP znajduje się funkcja nielinowa. Typem sieci MLP służącym do rozpoznawania obrazów jest splotowa sieć neuronowa inaczej znana jako sieć konwolucyjna. Tego typu sieć została przedstawiona w wcześniej wspomnianej pracy [17]. Autorzy użyli dwóch sieci typu MLP. Jednej dla znaków trójkątnych, a drugiej dla okrągłych. Wielkość wejścia sieci odpowiadała obrazowi o rozmiarach 30x30 piksel. Warstwa wyjściowa sieci posiadała 10 wyjść z czego 9 wyjść odpowiadało znakom. Ostatnie wyjście informowało, że znak nie został rozpoznany. Autorzy eksperymentalnie dobrali wielkość sieci z pośród 3 typów sieci. Najlepsze wyniki uzyskała 3 konfiguracja 30x30/15/5/10 z 98% skutecznością dla najlepiej rozpoznawalnych znaków. Rozpoznawanie na jednostce obliczeniowej PC486 33 MHz zajmuje 220 ms dla obrazu w rozdzielczości 256x256 co jest znakomitym wynikiem jak na rok 1996. Należy jednak zaznaczyć, że segmentacja w tej pracy była bardzo kosztowna czasowo, co nie pozwoliło algorytmowi działać w czasie rzeczywistym.

W innej pracy przedstawionej w 2003 roku [13] również użyto splotowych sieci neuronowych do rozpoznawania znaków drogowych. W tej pracy segmentacja obszaru występowania znaku przeprowadzona była za pomocą progowania oraz algorytmów genetycznych. Sieć była uczona za pomocą algorytmu wstecznej propagacji błędu (ang. Backpropagation Neural Network), według paradygmatu ART1 (ang. Adaptive Resonance Theory). Sieć posiada dwie warstwy ukryte. Dla całego algorytmu przedstawionego w pracy rozpoznanie obrazu wraz z klasyfikacją znaków zajmuje mniej niż 8.8 ms na procesorze AMD Duron 1 GHz. Biorąc jednak pod uwagę, że dla systemów TSR rozpoznanie w czasie rzeczywistym liczone jest od prędkości 0.2 s dla jednego obrazu wynik pracy jest bardzo dobry. W pracy [29] opracowano algorytm bazujący na sieci neuronowe propagacji wstecznej pozwalający na śledzenie obiektów. Na potrzeby algorytmu stworzono dwie sieci działające równolegle ze sobą. Jedna z nich stara się zidentyfikować obiekt na bazie koloru, a druga na bazie wykrytych krawędzi. Obiekt jest rozpoznawany i klasyfikowany a następnie śledzony. Czas rozpoznania obrazu wynosi 0.3 s. Sieć neuronowa uczona algorytmem wstecznej propagacji została również zaimplementowana w 2005 r. przez autorów artykułu[30]. Autorzy wspomnieli, że proces uczenia sieci algorytmem wstecznej propagacji jest kosztowny pod względem obliczeniowym oraz łatwo podczas szkolenia sieci wpaść w lokalne minimum co przysparza problemów. By zapobiec ostatniemu problemowi część autorów algorytmów starała się uczyć sieć metodami głębokiego uczenia DL (ang. Deep Learning). Warto zauważyć, że przy uczeniu sieci jakąkolwiek metodą ważne jest by niektóre obrazy uczące były w pewien sposób zaszumione oraz obrócone pod małym kontem. Taki zabieg powoduje większą odporność sieci. W pracy [31] szkolono dwie sieci wielowarstwowe metodą głębokiego uczenia. Jedna sieć była przeznaczona do segmentacji obrazów, a druga do klasyfikacji. Obie sieci posiadały podobną strukturę z wyjątkiem ostatniej warstwy. Przy testowaniu sieci przeznaczonej do segmentacji obrazu zauważono, że sieci splotowe są znacznie wydajniejsze, gdy zostają użyte w trybie przesuwanego okna. Poza tym część informacji z nakładających się regionów można użyć wielokrotnie co przyspiesza algorytm. Architektura sieci przedstawiona jest na Rysunek 8 Architektura sieci splotowej z rozgałęzieniem po 6 warstwie

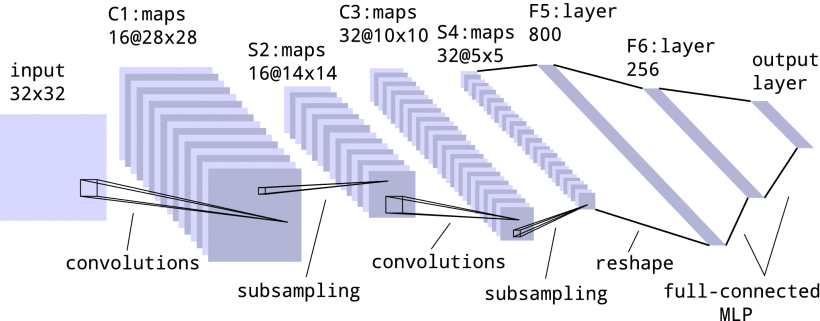


Rysunek Architektura sieci splotowej z rozgałęzieniem po 6 warstwie

Przy szkoleniu sieci niektóre próbki szkoleniowe były zniekształcone o ± 20˚ oraz dodano obrazy nie zawierające znaków drogowych. Łączenie użyto około 6600 obrazów do szkolenia i 3400 do testowania sieci. Do testów została użyta jednostka Intel Xeon E5-1620, dwoma kartami NVIDIA Tesla K20 G-PU i 32 GB pamięci. Sieć wykrywa znaki drogowe w ciągu 0.8 s z 88 % skutecznością.

#### Konwolucyjne sieci neuronowe

CNN (ang. Convolutional Neural Network) czyli splotowe sieci neuronowe, nazywa się takie sieci, których pierwsze warstwy dokonują splotu danych wejściowych. Takie sieci zostały wprowadzone by zapobiec problemowi przeuczenia się sieci. Splot może polegać na łączeniu informacji z kilku pikseli i przekazywaniu jej jako ujednolicona wartość. W pracy [16] zastosowano splotową sieć neuronową dla by sklasyfikować znaki drogowe. W tym celu wyszkolono trzy siec, każda dla innego typu znaku. Na wejście dostarczano obrazy w skali szarości i rozmiarze 32×32 pikseli. Architektura sieci przedstawiona jest na Rysunek 9



Rysunek Struktura sieci splotowej opisywanej pracy

Ponieważ obrazy są przechwytywane w różnych warunkach oświetleniowych i pogodowych, znaki tej samej klasy mogą znacząco różnić się od siebie. Aby zmniejszyć różnicę jasności użyto metody korelacji adaptacyjnego histogramu ograniczonego kontrastem (CLAHE), pozwalającej dostosować kontrast obrazów. Z powodu tego, że wykryte regiony mogą zawierać błędy pierwszego i drugiego rodzaju dodano klasę tła w trakcie uczenia. Każda z sieci zawierała dwie warstwy splotowe i dwie warstwy posiadające architekturę sieci MLP. Rozmiar obrazu wejściowego wynosi 32×32. Rozmiar filtru w obu warstwach splotowych wynosił 5×5. Po pierwszej warstwie splotu, zostało dodanych się 16 map powodując zmniejszenie rozmiaru wejścia drugiej warstwy splotowej najpierw do rozdzielczości 28×28, a następnie do 14×14. Algorytm po transformacjach obrazu i zmapowaniu dostarcza wektor o długości 800 znaków. Sieć jest zaprogramowana w języku C++ z wykorzystaniem biblioteki OPENMP. Wykorzystywane są również implementacje MSER, HOG i SVM z biblioteki OpenCV. Sieć neuronową wyszkolono za pomocą Torch7 i przepisano wyszkolone wartości do kodu C++. Algorytm został uruchomiony na komputerze z czterordzeniowym procesorem 3,7 GHz, poprawnie wykrywając z dokładnością 99.29 % znaków zakazu, 96.74 % znaków nakazu i 97.13 % znaków ostrzegawczych przy czasie wykrywania wynoszącym 0.162 s. Znaki do testowania wzięto z niemieckiego benchmarku GRSDB.

Architektura sieci splotowej przedstawionej w artykule [14] różni się od tradycyjnych sieci splotowych typem nieliniowości, ponieważ w węzłach sieci użyto funkcji przypominającą funkcję tangensa hiperbolicznego. Dodatkowo wykorzystano połączenia między niesąsiadującymi warstwami sieci. Tradycyjne sieci splotowe nie posiadają połączeń między warstwami niesąsiadującymi. W warstwie poolingu użyto różnych parametrów odpytywania dla danych dostarczanych z różnych warstw. Sieć została zaimplementowana z użyciem biblioteki EBLearn, oraz była również testowana na znakach pochodzących ze zbioru GTSRB. W przeciwieństwie do poprzedniej sieci ta sieć była uczona dla wszystkich zbiorów znaków jednocześnie. Co ciekawe sieć osiągnęła wynik 98.97 % dokładności wykrywania dla architektury, w której zostały wykorzystane informację o kolorach i 99.17 % dokładności dla obrazów w skali szarości. Jest to sprzeczne z ogólnym poglądem twierdzącym, że wykorzystanie kolorów powinno poprawić skuteczność algorytmu.

#### Extreme Learning Machine

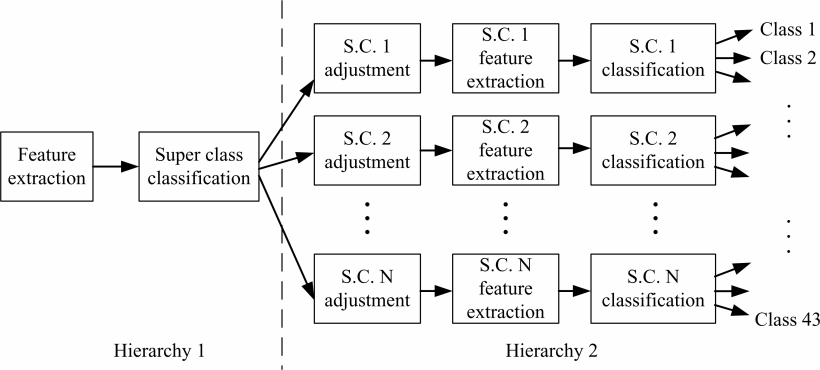
ELM (ang. Extreme Learning Machine) jest algorytmem uczenia się pojedynczych sieci ukrytych sieci neuronowej (SFNN). Pierwszą zaletą algorytmu ELM jest to, że wartości wejściowe między wejściem i warstwami ukrytymi są losowo przypisywane, co pozwala losowe mapowanie cech. Ponieważ wyszkolone są tylko warstwy wyjściowe między warstwami ukrytymi i wyjściowymi, nie jest wymagane strojenie za pomocą wstecznej propagacji warstwa po warstwie. Algorytm ELM może uzyskać optymalne i uogólnione rozwiązanie do rozpoznawania wieloklasowego. Dodatkowo, łatwo można rozszerzyć ELM na sieć wielowarstwową lub głęboką za pomocą techniki autoencoder. Metoda ELM została również wykorzystana do modelowania lokalnych pól recepcyjnych i wykorzystywana do uczenia na dużych zbiorach danych. W związku z tym zastosowanie ELM dla TSR może dać lepsze rozwiązanie. Co więcej, z powodu losowego przypisania wag wejściowych, algorytm ELM może zmniejszyć koszt obliczeniowy szkolenia. Ponieważ istnieje tylko jedna ukryta warstwa, szybkość obliczeniowa procesu rozpoznawania jest również szybka. Praca [22] stosuje jako klasyfikator pojedynczą sieć z ukrytą warstwą. W oparciu o algorytm ELM, połączenie pomiędzy warstwami wejściową i ukrytą realizuje mapowanie cech losowych, podczas gdy tylko wagi pomiędzy warstwami ukrytymi i wyjściowymi są wyszkolone. W rezultacie strojenie warstwa po warstwie nie jest wymagane. Tymczasem norma wag wyjściowych jest zawarta w funkcji kosztów. W związku z tym, Klasyfikator oparty na ELM może osiągnąć optymalne i uogólnione rozwiązanie dla wieloklatkowych TSR. Ponadto może zrównoważyć dokładność rozpoznawania i koszty obliczeniowe.

#### Liniowa analiza dyskryminacyjna

LDA (ang. Linear Discriminant Analysis) - liniowa analiza dyskryminacyjna jest to algorytm, którego używa się w uczeniu maszynowym do znalezienia liniowej kombinacji cech, które najlepiej rozróżniają dwie lub więcej klas obiektów lub zdarzeń. Wynikowe kombinacje są używane jako klasyfikator liniowy lub częściej, służą redukcji wymiarów do późniejszej klasyfikacji statystycznej. [here] (usunąć cały opis??)

### Maszyna wektorów nośnych do klasyfikacji

W artykule [19] zaproponowano innowacyjną metodę rozpoznawania znaków nazwaną hierarchiczną metodą wektorów nośnych. Algorytm zajmuje się wyłącznie klasyfikacją znaków drogowych bez etapu segmentacji i był uczony na zbiorze GTSRB. Algorytm działa w następujący sposób. Wpierw znak drogowy klasyfikowany jest przez SVM do jednej z trzech superklas. Następnie w zależności do jakiej klasy znak drogowy został zakwalifikowany zostaje użyta funkcja dostosowania perspektywy z różnymi parametrami. Po dostosowaniu perspektywy ponownie zostają użyte maszyny wektorów nośnych zdolne rozpoznać dokładnie co jest przedstawione na znaku. W roli wyjaśnienia jedna maszyna klasyfikuje znaki do podklas, a dla każdej podklasy wytrenowane są inne maszyny zdolne do odczytania znaku. Pierwsza maszyna SVM przydziela klasę znaku z pomocą algorytmu HOG. Dzięki rozpoznaniu superklasy znaku można wykorzystać informację o kolorze. Skorygowanie perspektywy jest kluczowym elementem pomagającym lepiej rozpoznać znak drogowy. Taka czynność eliminuje wszystkie translacje i rotację znaku co pozwala lepiej odczytać znak w następnym etapie. Schemat blokowy proponowanej metody znajduje się Rysunek 10.



Rysunek Schemat blokowy metody opartej na SVM

Algorytm dokonuje wstępnej ekstrakcji cech obrazu za pomocą wyszukania czerwonych pikseli oraz sprawdzeniu pikseli sąsiadujących z czerwonymi pikselami. Po wstępnym przetworzeniu obrazu następuje znalezienie okręgów na obrazie za pomocą radialnego detektora symetrii. Radialny detektor symetrii jest wariantem kołowej transformaty Hougha. Po wykryciu okręgów następuje wykrycie trójkątów, czyli znaków ostrzegawczych. Do wykrywania trójkątów na obrazie również użyta zostaje transformata Hougha wykrywająca linie. Następnie algorytm wykrywa linie przecinające się mogące stanowić kontury znaku. Na chwilę obecną algorytm zaproponowany w pracy z 43 klas znaków dostępnych w GTSRB rozróżnia znaki ostrzegawcze i zakazu będące okręgami i trójkątami z dokładnością 99.03 %. Znaki należące do innych klas są klasyfikowane za pomocą algorytmu SVM i HOG bez informacji o kolorze z dokładnością 99,94%, 98,89% i 99,90% dla znaków kolejno nakazu, ograniczeń prędkości i znaków pozostałych, co daje globalny wynik dokładności algorytmu na poziomie 99,52% z średnią szybkością 40 ms. Implementacja została wykonana za pomocą programy Matlab 2011b i na jednostce Core 13 3.3 GHz.

Praca [11] proponuje kompletny algorytm bazujący na SVM pozwalający zlokalizować i rozpoznać znak drogowy. Wcześniej zaznaczono, że segmentacja dokonywana jest w przestrzeni HSI, a regiony wykryte przez segmentację wstępnie są przydzielane do podklas za pomocą linowych maszyn SVM. Pozwala to wstępnie rozróżnić typ znaku. Rozpoznawanie realizowane jest przez maszyny SVM z jądrami Gaussa. Praca do treningu wykorzystuje bibliotekę LIBSVMS. Starano się użyć funkcji liniowej jako hiperpłaszczyzny, lecz nie zawsze było to możliwe. W trudnych przypadkach wprowadzono jądro Gaussa. Na wejście SVM podawano 31x31 pikseli, dlatego każdy obszar, gdzie zostały wykryte potencjalne znaki musiał być przeskalowany do tej rozdzielczości. Dla każdego koloru i kształtu podano od 20 do 100 próbek uczących. Przetwarzanie pojedynczej ramki o wielkości 720×576 pikseli za pomocą proponowanej metody zajmuje 1.77 s na procesorze Pentium 4-M 2,2 GHz. Testy prowadzone były na autorskim benchmarku znaków obowiązujących w Hiszpanii.

### Drzewo k-wymiarowe

Praca [32] jest przykładem artykułu starającym się rozważyć przydatność drzew wielowymiarowych K-d (ang. K-Dimensional Tree) do rozróżniania znaków drogowych. Drzewo K-d jest wariantem drzewa binarnego opartego o wyszukiwanie najbliższego sąsiada. Zaletą drzew K-d jest łatwość budowy i aktualizacji oraz szybkość ich przeszukiwania. Dodatkowo drzewa dobrze radzą sobie z niezrównoważonymi zbiorami danych w przeciwieństwie do sieci neuronowych, które wymagają dobrego dostrojenia parametrów. W przypadku pracy drzewo K-d składa się z informacji dostarczonych od czterech transformat HOG na obrazie, każdej z innymi parametrami. Do wyszukiwania najbliższego sąsiada użyto algorytmu Best Bin First. Na znakach dostarczonych przez zbiór danych GTSRB drzewo K-d bazujące na deskryptorze HOG przedstawione w pracy uzyskało poprawność rozpoznania na poziomie 92,9%.

### Lasy losowe

W tej samej pracy co drzewa K-d [32] testowano również lasy losowe RM (ang. Random Forests) do problemu klasyfikacji znaków drogowych. Klasyfikacja obrazu przez las losowy polega na zebraniu odpowiedzi z kilku drzew decyzyjnych przez głosowanie. Lasy losowe również jak drzewa D-d dobrze radzą sobie z klasyfikacją na zbiorze niesymetrycznym niekiedy przewyższając niekiedy algorytmy SVM. Jak i w drzewie K-d opisywanym powyżej obraz wejściowy poddawany był 4 transformatą HOG. Dla lasu losowego testowanego na tym samym zbiorze danych GTSRB poprawność klasyfikacji została poprawiona do 97,2%.

## Podsumowanie

Praca miała na celu zbadanie w jaki sposób zastosowanie wielu kamer może polepszyć wykrywanie znaków drogowych. Pierwszym etapem przeglądu literaturowego było opisanie możliwych konfiguracji kamer np. poprzez zastosowanie kamer z IR lub łączenia klatek filmu w panoramę. Do łączenia obrazów w panoramę zostało wyszukanych kilka metod, z czego najskuteczniejsze okazały się metody bazujące na punktach charakterystycznych, które zostały opisane. Są one szybkie w działaniu oraz odporne na małe rotacje dwóch obrazów. Należy również zaznaczyć, że problem tworzenia panoramy na potrzeby pracy niesie ze sobą wiele udogodnień. Z dość dużą dokładnością znana jest translacja pozioma obu obrazów. Translacja pionowa praktycznie nie występuje, co prowadzi do wniosku znany jest obszar, w którym obrazy z kamer się nakładają. Umożliwia to wykonywanie operacji dopasowania tylko na wspólnym obszarze rejestrowanym przez obie kamery co pozwoli zmniejszyć ilość obliczeń. Przez wprowadzone ograniczenia metody klasyczne nie są od skazane na niepowodzenie. Metody bazujące na punktach charakterystycznych wydają się przynosić najbardziej obiecujące wyniki.

Podsumowując zagadnienie segmentacji należy stwierdzić, że podejścia oparte na kolorze są nadal stosowane ze względu na ich szybkość. Ich zawodność jest natomiast poprawiana dodatkowymi algorytmami działającymi poprzez wykrywanie kształtu. Podczas progowania najczęściej zostają odnalezione grupy pikseli nie powiązane ze sobą. By zlokalizować obiekty odnalezione przez progowanie należy w jakiś sposób połączyć piksele ze sobą i je oznaczyć. Dość prostą i skuteczną metodę przedstawiono w pracy [9] gdzie do każdego odnalezionego piksela zostaje przypisana etykieta. Algorytm etykietowania wprowadza osobny bufor umożliwiający łączenie regionów z różnymi etykietami. Podejścia oparte na teksturze (HOG, LPB, ACF) z dużą dokładnością potrafią wyłonić obrys znaku drogowego, co na etapie identyfikowania może posłużyć do szkolenia klasyfikatora. Niestety takie przetwarzanie na całym obrazie jest zbyt kosztowne obliczeniowo, dlatego zawęża się obszar stosując ROI. Pozwala to przyspieszyć obliczenia i umożliwia przetwarzanie obrazu w czasie rzeczywistym. Obszar ROI dotyczy głównie miejsc środkowych i po prawej stronie. Niektóre algorytmy jak ten zaproponowany w pracy [33] używają funkcji ACF z klasyfikatorem AdaBoost. Algorytm ACF został również użyty w pracy [34] w połączeniu z informacją o kolorach i gradientach oraz algorytmem HOG i uzyskał dobre wyniki. Oprócz powyższych podejść, splotowa sieć neuronowa (CNN) jest stosowana do wykrywania znaków drogowych i osiąga doskonałe wyniki [14]. Warto również wspomnieć, że dla lepszej wstępnej filtracji obrazu można zastosować również Filtry Gabora lecz obecnie używanie takiego algorytmu jest nieefektywne ze względu na korzyści i czas obliczeń. Wartym wspomnienia podejściem detekcji znaków jest tworzenie mapy cieplnej obszarów występowania znaków [35]. Taka mapa działa na zasadzie algorytmów biologicznych. Warto również dodać, że część obliczeń wykorzystanych w etapie tworzenia panoramy może posłużyć w etapie segmentacji obiektów.

Najpowszechniej stosowanym podejściem do klasyfikacji obiektów są splotowe sieci neuronowe. Czas klasyfikacji odnajdywania obiektów przez sztuczne sieci neuronowe balansuje na granicy pracy w czasie rzeczywistym, dlatego niektórzy autorzy starają się przyspieszyć wykrywanie znaków drogowych łącząc sieci CNN z innymi algorytmami np. w pracy [16] algorytmem SVM. Drzewa wielowymiarowe oraz lasy losowe chodź osiągają znakomite wyniki niestety okazuje się, że są zbyt wole by pracować w czasie rzeczywistym.

# Program

Podczas projektowania programu musiano dokonać wielu wyborów. Pierwszym z nich był wybór w jaki sposób należy rozszerzyć system o zastosowanie większej ilości kamer. Ze względu na koszty kamery podczerwonej zrezygnowano z jej zastosowania. Zdecydowano rozszerzyć system o zastosowanie trzech jednakowych kamer pochodzących obejmujących obraz wokół samochodu. Prototyp systemu był zbudowany na trzech kamerach. Dwie z nich był umiejscowione na daszku przeciwsłonecznym kierowcy i pasażera i obejmowały obszar bezpośrednio przed pojazdem. Trzecia była skierowana w taki sposób by symulowała widok pasażera patrzącego w prawą stronę. Rysunek 16 przedstawia widoki dla każdej z kamer

|  |
| --- |
| Obraz zawierający zewnętrzne, budynek, droga, ulica  Opis wygenerowany automatycznieObraz zawierający droga, zewnętrzne, ulica, budynek  Opis wygenerowany automatycznieObraz zawierający niebo, zewnętrzne, podłoże, budynek  Opis wygenerowany automatycznie |

Rysunek Widok z trzech kamer wybrany dla systemu

Konfiguracja kamer pozwalała zarówno na stworzenie z nich panoramy jak i zastosowanie przetwarzania równoległego. Wpierw podjęto próbę stworzenia obrazu panoramicznego. Zdecydowano, że najlepszym rozwiązaniem będzie stworzenie takiego obrazu z pomocą wcześniej opisywanych metod bazujących na punktach kluczowych. Pierwszym algorytmem był algorytm SIFT (…)

Rozwiązaniem, które zostało wybrane dla tej pracy jest przetwarzanie równoległe każdego z obrazów dostarczonych przez kamery tego samego typu ustawionych co pewne przesunięcie. Takie rozwiązanie zostało wybrane, tworzenie panoramy ze zdjęć jest zbyt kosztowne czasowo.

## Wyszukiwanie znaków na obrazie

Segmentacja obrazu w postaci wyszukiwania obszarów mogących zawierać znaki była jednym z ważniejszych elementów całej pracy. Realizacje tego zadania podzielono na kilka etapów. Każdy z etapów został przedstawiony poniżej. Etap przedstawiono na przykładzie obrazu pochodzącym z kamery po lewej stronie. Identyczny proces przeprowadzany był dla obrazu dostarczonego z kamery środkowej i po prawej stronie. Wszystkie operacje na obrazie przeprowadzane były z udziałem biblioteki OpenCV.

### Segmentacja

Pierwszą operacją wykonywaną na obrazie była jego konwertowanie do rozdzielczości 640 x 360 pikseli. Taki obraz przedstawia Rysunek 17.

Obraz zawierający zewnętrzne, budynek, droga, ulica

Opis wygenerowany automatycznie

Rysunek Oryginalny obraz wejściowy

Następnie z obrazu zostawiono rejon, gdzie znaki mogły się znaleźć z największym prawdopodobieństwie, a resztę ucięto. Taki obszar nazwany jest regionem zainteresowań ROI. Powoduje to mniejszą ilość obliczeń, a co za tym idzie przyspiesza działanie algorytmu. Powoduje też wykrycie mniejszej ilości fałszywych regionów. Obszar ROI przedstawia Rysunek 18.

Obraz zawierający zewnętrzne, budynek, drzewo, droga

Opis wygenerowany automatycznie

Rysunek Obraz po wyznaczeniu ROI

Następnym krokiem było przetworzenie obrazu poprzez wyrównanie histogramu osobno dla każdej składowej piksela. Osobno dla czerwonej, zielonej i niebieskiej składowej. Efekt takiego działania został zaprezentowany na Rysunek 19.

Obraz zawierający zewnętrzne, budynek, droga, ulica

Opis wygenerowany automatycznie

Rysunek Obraz po wyrównaniu histogramów

Następnie obraz konwertowano do przestrzeni HSV. Przestrzeń została wybrana ze względu na swoją odporność na różnicę w jasności obrazu. W tym miejscu obraz wejściowy został skopiowany i przetwarzany przez trzy niezależne detektory. Detektory podzielone były ze względu na kolor znaków. Znajdowano znaki koloru niebieskiego, czerwonego i żółtego. Pierwszym krokiem każdego z detektorów było progowanie. Wartości progów dobrano eksperymentalnie po przeprowadzeniu parudziesięciu testów znaków z pomiędzy dla koloru od 0 do 180. Dla nasycenia i wartości barwy od 0 do 255. Przedstawia je Tabela 1.

|  |  |  |  |
| --- | --- | --- | --- |
| Kolor | Niebieski | Czerwony | Żółty |
| Minimalny kąt koloru | 92 | 129 | 14 |
| Maksymalny kąt koloru | 126 | 180 | 31 |
| Minimalne nasycenie | 197 | 101 | 116 |
| Maksymalne nasycenie | 255 | 255 | 255 |
| Minimalna wartość barwy | 41 | 30 | 112 |
| Maksymalna wartość barwy | 255 | 255 | 255 |

Tabela Wartość parametrów progowania dla trzech kolorów

Rysunek 20 przedstawia efekt progowania dla każdego z kolorów.

|  |
| --- |
| 1. – Progowanie w kolorze niebieskim |
| 1. – Progowanie w kolorze czerwonym |
| 1. – Progowanie w kolorze żółtym |

Rysunek Progowanie wykonane dla każdego z kolorów

Można zauważyć, że progowanie wyłoniło obrys znaków drogowych, lecz również nieco obiektów, które nimi nie są. Problemem, z którym należało się zmierzyć było odfiltrowanie obszarów najmniej prawdopodobnych. Do tego zadania zdecydowano się zastosować operacje morfologiczne na obrazie. Inne operacje były stosowane dla koloru niebieskiego, czerwonego i żółtego. Filtry zostały przedstawia Tabela 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Kolor | Niebieski | | Czerwony | | Żółty | |
| Operacje | Dylatacja | 1 | Domknięcie | 1 | Otwarcie | 1 |
| Domknięcie | 5 | Domknięcie | 2 |

Tabela Operacje morfologiczne użyte na obrazie binarnym

Jądrem filtru była siatka 3x3 piksele. Wybór parametrów został dobrany tak samo jak parametry progowania na podstawie badań i doświadczeń. Zauważono, że znaki niebieskie posiadają duży wycinek obszaru białego, czerwonego bądź czarnego, który nie zostaje uwidoczniony podczas progowania. Powoduje to, że znak na obrazie podzielony jest na fragmenty. By połączyć elementy znaku ze sobą zastosowano filtry przedstawione w Tabela 1. Znaki koloru czerwonego wykrywane były za pomocą detekcji okręgu na obrazie. Jak można dostrzec na Rysunek 20 obrazie (b) okrąg jest dobrze widoczny po samym progowaniu oraz znikoma część innych obiektów posiadała takie właściwości, dlatego zdecydowano się tylko użyć filtra pojedynczego domknięcia. Dla znaków drogowych w kolorze żółtym zdecydowano się zastosować operacje otwarcia i podwójnego domknięcia. Za takim wyborem przemawiały dwa argumenty. Pierwszym z nich jest, że znaki informacyjne nie posiadają obszaru białego na którym umiejscawia się symbol, a jedynie sam symbol znaku. Pozwala to usunąć pojedyncze piksele nie tracąc kształtu znalezionego znaku. Operacje domknięcia pogrubiają obrys znaku by był bardziej widoczny. Efekty nałożonych filtrów zaprezentowane są na Rysunek 22.

|  |
| --- |
| 1. – progowanie niebieske po operacjach morfologicznych |
| 1. – progowanie czerwone po operacjach morfologicznych |
| (c) – progowanie żółte po operacjach morfologicznych |

Rysunek Obrazy po zastosowaniu operacji morfologicznych z Tabela 1

Na obrazach binarnych poddanych operacjami morfologicznymi wyszukiwano obszary mogące zawierać znak. Dla obrazów progowanych w kolorze niebieskim i żółtym zastosowano wbudowany funkcję Blob Detector, czyli detektor plam. Detektor ten przyjmował zestaw danych, które pozwalały wyznaczać parametry wyszukiwanych obiektów. Parametry te przedstawia Tabela 3. Tak jak parametry progowania zostały one dobrane eksperymentalnie.

|  |  |  |
| --- | --- | --- |
| Kolor | Niebieski | Żółty |
| Minimalna wielkość | 68 | 40 |
| Maksymalna wielkość | 897 | 1000 |
| Minimalny próg jasności | 0 | 0 |
| Maksymalny próg jasności | 255 | 255 |
| Minimalna kolistość | 0,38 | 0,45 |
| Maksymalna kolistość | 1 | 1 |
| Minimalne wydłużenie | 0,2 | 0,6 |
| Maksymalne wydłużenie | 1 | 1 |
| Minimalna wklęsłość | 0,61 | 0,73 |
| Maksymalna wklęsłość | 1 | 1 |

Tabela Parametry Blob Detector'a

Wielkość opisywana jest jako ilość pikseli przynależąca do danego obiektu. Minimalny i maksymalny próg jasności brany był z całego dostępnego zakresu od 0 do 255. Kolistość obiektu wskazuje jakie wartości kątów wklęsłych posiada. Czym większa wartość kątów tym obiekt jest bardziej kolisty. Ostatnimi parametrami wejściowymi było wydłużenie i wklęsłość obiektu. Prócz dwóch pierwszych parametrów były one brane z zakresu od 0 do 1. Przykłady obiektów z różnymi parametrami przedstawia Rysunek 23

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
| Wielkość | Jasność | Kolistość | Wydłużenie | Wypukłość |

Rysunek Przykłady obiektów z różnymi parametrami

Jak wcześniej wspomniano elementy czerwone w znakach zakazu są okręgami. Zdecydowano więc by wyszukiwać je na obrazie nie poprzez funkcję Blob Detector lecz za pomocą transformaty Hougha przystosowanej do wyszukania okręgów. Parametry wejściowe przedstawione są w Tabela 4.

|  |  |
| --- | --- |
| Kolor | Czerwony |
| Metoda | Gradientowa |
| Wymiar akumulatora (dp) | 1 |
| Minimalny dystans między okręgami | 15 |
| Próg krawędzi | 1 |
| Próg dla centrów okręgu | 15 |
| Minimalny promień | 6 |
| Maksymalny promień | 18 |

Tabela Parametry transformaty Hougha dla okręgów

Metoda jaką była liczona transformata to jedyna zaimplementowana metoda w OpecCV – gradientowa (Hough Gradient). Następnym parametrem był wymiar akumulatora liczony jako odwrotny stosunek rozdzielczości akumulatora do rozdzielczości obrazu. Na przykład, jeśli dp = 1, akumulator miał taką samą rozdzielczość jak obraz wejściowy. Jeśli dp = 2, akumulator miał połowę szerokości i wysokości. Wybrano by akumulator miał taki sam wymiar jak obraz wejściowy. Minimalna wartość pomiędzy dwoma wykrytymi okręgami wynosiła 15 pikseli. Spowodowało to, że wykryte okręgi mogące symbolizować znaki nie nakładały się na siebie. Następnie należało wybrać najwyższy próg przekazanych do detektora krawędzi. Dolny próg był dwa razy mniejszy od górnego. Następnie należało zadać próg akumulatora dla centrów okręgu na etapie wykrywania. Im próg był mniejszy, tym więcej fałszywych okręgów mogło być wykrytych. Koła, z większymi wartościami akumulatora zostały zwracane jako pierwsze, dlatego uznano, że próg ten powinien mieć dużą wartość. Ostatnimi parametrami był zakres długości promienia dla wykrytych okręgów.

Dla obiektów wykrytych przez metodę wykrycia plam jak i przez transformatę Hougha kopiowano wycinek nieprzetworzonego obrazu odpowiadający miejscu wykrycia. W ten sposób tworzono w obrazy mogące być przetwarzane w części klasyfikacji. Przykłady tych obrazów przedstawia Rysunek 24.

|  |  |  |
| --- | --- | --- |
| Niebieskie | Czerwone | Żółte |
| Obraz zawierający wewnątrz  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, wewnątrz  Opis wygenerowany automatycznieObraz zawierający rozmyty  Opis wygenerowany automatycznieObraz zawierający zewnętrzne, transport, rozmyty, samochód  Opis wygenerowany automatycznieObraz zawierający drzewo, rozmyty  Opis wygenerowany automatycznieObraz zawierający osoba  Opis wygenerowany automatycznieObraz zawierający rozmyty  Opis wygenerowany automatycznieObraz zawierający rozmyty, zewnętrzne  Opis wygenerowany automatycznieObraz zawierający rozmyty, budynek  Opis wygenerowany automatycznieObraz zawierający rozmyty  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, niebo, zewnętrzne  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, wewnątrz  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, zwierzę, osoba  Opis wygenerowany automatycznieObraz zawierający rozmyty, budynek  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, transport, zdjęcie  Opis wygenerowany automatycznieObraz zawierający wewnątrz  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, wewnątrz  Opis wygenerowany automatycznieObraz zawierający budynek  Opis wygenerowany automatycznieObraz zawierający sprzęt elektroniczny  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, zdjęcie  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, wewnątrz  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, wewnątrz, podłoże  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, osoba, zewnętrzne  Opis wygenerowany automatycznieObraz zawierający wewnątrz, rozmyty, zdjęcie, ściana  Opis wygenerowany automatycznieObraz zawierający obiekt  Opis wygenerowany automatycznieObraz zawierający obiekt  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, zdjęcie  Opis wygenerowany automatycznieObraz zawierający rozmyty  Opis wygenerowany automatycznieObraz zawierający zewnętrzne  Opis wygenerowany automatycznieObraz zawierający rozmyty, zwierzę  Opis wygenerowany automatycznieObraz zawierający tenis, rakieta, gracz, piłka  Opis wygenerowany automatycznieObraz zawierający transport  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, trawa, budynek  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, zdjęcie, siedzi  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, zdjęcie, budynek  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, butelka, siedzi  Opis wygenerowany automatycznieObraz zawierający napoje, żywność  Opis wygenerowany automatycznieObraz zawierający napoje, żywność  Opis wygenerowany automatycznieObraz zawierający rozmyty  Opis wygenerowany automatycznieObraz zawierający zewnętrzne  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, zdjęcie  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, monitor, góra  Opis wygenerowany automatycznieObraz zawierający rozmyty  Opis wygenerowany automatycznieObraz zawierający samochód, zewnętrzne, transport  Opis wygenerowany automatycznieObraz zawierający wewnątrz  Opis wygenerowany automatycznieObraz zawierający urządzenie  Opis wygenerowany automatycznieObraz zawierający osoba, odzież, stojące, zdjęcie  Opis wygenerowany automatycznieObraz zawierający zewnętrzne  Opis wygenerowany automatycznieObraz zawierający wewnątrz, rozmyty, urządzenie, siedzi  Opis wygenerowany automatycznie | Obraz zawierający roślina, drzewo, okno, budynek  Opis wygenerowany automatycznieObraz zawierający czerwony  Opis wygenerowany automatycznieObraz zawierający obiekt, wskaźnik  Opis wygenerowany automatycznieObraz zawierający wewnątrz, zdjęcie  Opis wygenerowany automatycznie | Obraz zawierający rozmyty, pociąg  Opis wygenerowany automatycznieObraz zawierający rozmyty, wewnątrz, zdjęcie  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, wewnątrz, ściana  Opis wygenerowany automatycznieObraz zawierający zawody lekkoatletyczne  Opis wygenerowany automatycznieObraz zawierający lalka  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie, wewnątrz  Opis wygenerowany automatycznieObraz zawierający rozmyty, zdjęcie  Opis wygenerowany automatycznieObraz zawierający rozmyty  Opis wygenerowany automatycznie |

Rysunek Przykłady wykrytych obiektów podczas segmentacji.

Na Rysunek 24 można dostrzec, że proponowany algorytm segmentacji jest w stanie wykryć znaki drogowe. Dokładne parametry skuteczności wykrytych regionów przedstawione zostały w akapicie Przygotowanie środowiska badawczego i plan badań. Tak przygotowanie obrazy zostały przekazywane do algorytmu odpowiedzialnego za klasyfikację.

- opisać operacje morfilogicznie

- opisać rozciągnięcie histogramu

### Klasyfikacja

# Propozycja rozwinięcia/konstrukcji metody

Użycie algorytmu rozmytego do dobrania parametrów

# Przygotowanie środowiska badawczego i plan badań

# Podsumowanie

# Bibliografia

|  |  |
| --- | --- |
| [1] | Y. Aoyagi i T. Asakura, „A study on traffic sign recognition in scene image using genetic algorithms and neural networks,” w *International Conference on Industrial Electronics, Control, and Instrumentation*, Taipei, Taiwan, Taiwan, 9 sierpień 1996. |
| [2] | Q. Hu, S. Paisitkriangkrai, C. Shen, A. v. d. Hengel i F. Porikli, „Fast Detection of Multiple Objects in Traffic Scenes With a Common Detection Framework,” *IEEE Transactions on Intelligent Transportation Systems,* tom 17, nr 7, pp. 1002-1024, kwieceń 2016. |
| [3] | A. Mogelmose, M. M. Trivedi i T. B. Moeslund, „Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems: Perspectives and Survey,” *Transactions on Intelligent Transportation Systems,* tom 13, nr 4, pp. 1484-1497, grudzień 2012. |
| [4] | B. Zitová i J. Flusser, „Image registration methods, a survey. Image and Vision Computing,” *Image and Vision Computing,* tom 21, nr 11, p. 977–1000, październik 2003. |
| [5] | F. Maes, A. Collignon, D. Vandermeulen, G. Marchal i P. Suetens, „Multimodality image registration by maximization of mutual information,” *Transactions on Medical Imaging,* tom 16, nr 2, p. 187–198, kwiecień 1997. |
| [6] | M. Kondej, B. Putz i M. Bartyś, „Szybki algorytm dopasowania obrazów dla potrzeb fuzji w czasie rzeczywistym,” 25 listopad 2010. [Online]. Available: https://automatykaonline.pl/Artykuly/Sterowanie/szybki-algorytm-dopasowania-obrazow-dla-potrzeb-fuzji-w-czasie-rzeczywistym. [Data uzyskania dostępu: 11 03 2019]. |
| [7] | J. Heather i M. Smith, „New adaptive algorithms for real-time registration and fusion of multimodal imagery,” 5 maj 2010. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7689/768908/New-adaptive-algorithms-for-real-time-registration-and-fusion-of/10.1117/12.850029.short?SSO=1. [Data uzyskania dostępu: 15 marzec 2019]. |
| [8] | D. G. Lowe, „Object Recognition from Local Scale-Invariant Features,” University of British Columbia, Vancouver, B.C., Canada, 1999. |
| [9] | E. Lee, S.-S. Lee, Y. Hwang i S.-J. Jang, „Hardware implementation of fast traffic sign recognition for intelligent vehicle system,” w *2016 International SoC Design Conference (ISOCC)*, Jeju, South Korea, 23-26 październik 2016. |
| [10] | D. Deguchi, M. Shirasuna, K. Doman, I. Ide i H. Murase, „Intelligent traffic sign detector: Adaptive learning based on online gathering of training samples,” w *2011 IEEE Intelligent Vehicles Symposium (IV)*, Baden-Baden, Germany, 05 lipiec 2011. |
| [11] | S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno i F. Lopez-Ferreras, „Road-Sign Detection and Recognition Based on Support Vector Machines,” *IEEE Transactions on Intelligent Transportation Systems,* tom 8, nr 2, pp. 264 - 278, 04 czerwiec 2007. |
| [12] | H. Liu, D. Liu i J. Xin, „Real-time recognition of road traffic sign in motion image based on genetic algorithm,” w *Proceedings. International Conference on Machine Learning and Cybernetics*, Beijing, China, 4-5 grudzień 2002. |
| [13] | A. d. l. Escalera, J. M. Armingol i M. Mata, „Traffic sign recognition and analysis for intelligent vehicles,” *Image and Vision Computing,* tom 21, nr 3, pp. 247-258, marzec 2003. |
| [14] | P. Sermanet i Y. LeCun, „Traffic sign recognition with multi-scale Convolutional Networks,” w *The 2011 International Joint Conference on Neural Networks*, San Jose, CA, USA, 21 lipiec 2011. |
| [15] | Y. Freund i R. E. Schapire, „Experiments with a New Boosting Algorithm,” w *Proceedings of the Thirteenth International Conference*, New Jersey, 1996. |
| [16] | Y. Yang, H. Luo, H. Xu i F. Wu, „Towards Real-Time Traffic Sign Detection and Classification,” *IEEE Transactions on Intelligent Transportation Systems,* tom 17, nr 7, pp. 2022 - 2031, czerwiec 2016. |
| [17] | A. d. l. Escalera, L. E. Moreno, M. A. Salichs i M. Armingol, „Road Traffic Sign Detection and Classification,” *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS,* tom 44, nr 6, pp. 848 - 859, grudzień 1997. |
| [18] | R. Duda i P. Hart, „Use of the Hough Transformation To Detect Lines and Pictures,” Stanford Research Institute, Menlo Park, California, 1972. |
| [19] | G. Wang, G. Ren, Z. Wu, Y. Zhao i L. Jiang, „A hierarchical method for traffic sign classification with support vector machines,” w *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA, 4-9 sierpień 2013. |
| [20] | I. M. Creusen, R. G. Wijnhoven, E. Herbschleb i P. D. With, „Color exploitation in hog-based traffic sign detection,” w *International Conference on Image Processing*, Hong Kong, 26-29 wrzesień 2010. |
| [21] | P. Sermanet, D. Eigen, Z. Xiang, M. Michael, F. Rob i L. Yann, „Integrated recognition localization and detection using convolutional networks,” Courant Institute of Mathematical Sciences, New York, NY, USA, 21 grudzień 2013. |
| [22] | Z. Huang, Y. Yu, J. Gu i H. Liu, „An Efficient Method for Traffic Sign Recognition Based on Extreme Learning Machine,” *IEEE Transactions on Cybernetics,* tom 47, nr 4, pp. 920 - 933, 14 marzec 2016. |
| [23] | T. Ojala, M. Pietikäinen i D. Harwood, „Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,” w *Proceedings of 12th International Conference on Pattern Recognition*, Jerusalem, Israel, Israel, 9-13 październik 1994. |
| [24] | P. Viola i M. Jones, „Robust Real-time Object Detection,” w *Second International Workshop on Statistical and Computational Theories of Vision*, Vancouver, Canada, 13 sierpień 2001. |
| [25] | C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer i T. Koehler, „A system for traffic sign detection, tracking, and recognition using color, shape, and motion information,” w *IEEE Proceedings. Intelligent Vehicles Symposium*, Las Vegas, NV, USA, USA, 6-8 czerwiec 2005. |
| [26] | N. Barnes i A. Zelinsky, „Real-time radial symmetry for speed sign detection,” w *IEEE Intelligent Vehicles Symposium*, Parma, Italy, Italy, 14-17 czerwiec 2004. |
| [27] | G. Loy i N. Barnes, „Fast shape-based road sign detection for a driver assistance system,” w *International Conference on Intelligent Robots and Systems*, Sendai, Japan, 28 wrzesień 2004. |
| [28] | T. Szymczyk, „Metoda dopasowania wzorców w rozpoznawaniu obrazów – ograniczenia, problemy i modyfikacje,” w *Automatyka*, Lublin, Wydział Elektrotechniki i Informatyki, Politechnika Lubelska, 2008, pp. 449 - 462. |
| [29] | C.-Y. Fang, S.-W. Chen i C.-S. Fuh, „Road sign detection and tracking,” *Transactions on Vehicular Technology,* tom 52, nr 5, pp. 1329 - 1341, 23 wrzesień 2003. |
| [30] | P. S. Miguel i A. R. Alastair, „Using self-organising maps in the detection and recognition of road signs,” *Image and Vision Computing,* tom 6, nr 27, pp. 673-683, maj 2009. |
| [31] | Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li i S. Hu, „Traffic-Sign Detection and Classification in the Wild,” w *2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27-30 czerwiec 2016. |
| [32] | F. Zaklouta, B. Stanciulescu i O. Hamdoun, „Traffic sign classification using K-d trees and random forests,” w *Proceedings of International Joint Conference of Neural Networks*, San Jose, California, USA, 31 lipiec 2011. |
| [33] | M. Mathias, R. Timofte, R. Benenson i L. V. Gool, „Traffic sign recognition-how far are we from the solution?,” w *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, USA, 4-9 sierpień 2013. |
| [34] | P. Dollár, Z. Tu, P. Perona i S. Belongie, „Integral Channel Features,” Dept. of Electrical Engineering, Los Angeles, Ca, USA, 2009. |
| [35] | R. Kastner, T. Michalke, T. Burbach, J. Fritsch i C. Goerick, „Attention-based traffic sign recognition with an array of weak classifiers,” w *2010 IEEE Intelligent Vehicles Symposium*, San Diego, CA, USA, 21-24 czerwiec 2010. |

Zagadnienie rozpoznania obrazu w większości przypadków realizowane jest w trzech krokach. Pierwszym krokiem jest przetwarzanie wstępne. Podczas takiego przetwarzania następuje np. akwizycja obrazu. W ogólności przetwarzanie-nisko-poziomowe ma na celu poprawę jakości obrazu poprzez eliminację zakłóceń, poprawę kontrastu, filtrację itp. Kolejny krok nazwany przetwarzaniem średniego poziomu i dotyczy segmentacji obrazu. Podczas segmentacji wydzielone zostają cechy obrazu. Przykładowo znajduje się punkty kluczowe lub obszary, w których mogą wystąpić obiekty. W pracy etap wstępnego przetwarzania omówiony został wraz z etapem segmentacji. Po segmentacji następuje próba rozpoznania i klasyfikacji obiektu. Bardziej zaawansowane metody pracujące na filmach posiadają również etap śledzenia wykrytych obiektów. Ponieważ każdy etapów przetwarzania obrazów można traktować jako oddzielny problem, poniżej w osobnych podrozdziałach zostały przedstawione metody rozwiązania każdego zagadnienia.