

## DETEKCJA KOLEJOWYCH SYGNALIZATORÓW ŚWIETLNYCH W OBRAZACH PRZY UŻYCIU DESKRYPTORÓW HOG I SIECI NEURONOWYCH SVM

W lipcu 2006 r. Navneet Dalal opublikował rezultaty badań detektora wykorzystującego sieci neuronowe SVM oraz deskryptory obrazów HOG (ang. Histogram of Oriented Gradients), który użyty został do wykrywania sylwetek ludzkich w obrazach i video. W tym artykule podobne badania zostały przeprowadzone dla detektora kolejowych sygnalizatorów świetlnych. W środowisku Matlab zaimplementowany został system TSR (ang. Traffic Sign Recognition) oraz przeprowadzona została jego analiza dla różnych parametrów deskryptorów HOG.

Pomimo, że system TSR ma takie same parametry jak system do detekcji sylwetek ludzkich, rezultaty badań tych systemów różnią się. Dobre rezultaty osiągnięto jedynie dla wybranej liczby poziomów kwantyzacji w deskryptorach HOG, podczas gdy w badaniach detektora ludzi uzyskiwano tym mniejsze wartości współczynnika strat (ang. miss rate), im większa była liczba poziomów kwantyzacji. Zwiększenie liczby obrazów w zbiorze uczącym klasyfikatora z 444 do 16000 poprawiło znacznie parametry detekcji. Niestety w badaniach systemu TSR uzyskano wysokie wartości współczynnika FPPW (ang. False Positives Per Window) w porównaniu do badań detektora sylwetek ludzkich. Może być to spowodowane brakiem normalizacji blokowej lub inną metodyką normalizacji w funkcji obliczającej deskryptor w programie Matlab. Wysokie wartości współczynnika FPPW mogą także świadczyć o tym, że rodzaj deskryptora lub metodyka detekcji są nieodpowiednie dla kolejowych sygnalizatorów świetlnych.

### WSTĘP

Wśród licznego zbioru systemów bezpieczeństwa w transporcie możemy wyróżnić systemy ostrzegania kierowców przed znakami drogowymi. Są to systemy TSR (ang. traffic sign recognition) [3]. Zazwyczaj są one stosowane w transporcie samochodowym. Znajdują one zastosowanie w utrzymaniu dróg (weryfikacji stanu sygnalizatorów świetlnych oraz znaków drogowych), w systemach ostrzegających kierowców o ograniczeniach prędkości oraz o znakach drogowych, a także w samochodach inteligentnych. W niniejszym artykule podjęto próbę zastosowania systemu TSR wykorzystującego sieci neuronowe SVM oraz deskryptory obrazów HOG (ang. Histogram of Oriented Gradients) w detekcji kolejowych sygnalizatorów świetlnych. Wadą systemów TSR jest ich niska sprawność. Wg [3] większość systemów wykrywa niewiele ponad 90% docelowych obiektów. W publikacji [1], w której badano system wykrywania sylwetek ludzkich (działający podobnie jak systemy TSR), dla pewnych konfiguracji detektora wykrywano 70% znajdujących się w obrazach osób. Zaletą tych systemów jest ich niski koszt.

Systemy, które ostrzegają kierowcę lub maszynistę komunikując się z nim na drodze radiowej lub przewodowej wymagają obecności układów elektronicznych w każdym z wykrywanych obiektów (przejazdy kolejowe, sygnalizatory świetlne). Systemy bezpieczeństwa oparte o komunikację radiową lub przewodową z kierowcą (lub maszynistą), jak również ich utrzymanie są kosztowne. W przypadku transportu samochodowego liczba znaków drogowych uniemożliwia stosowanie systemów ich radiowej identyfikacji. W systemach TSR nie ma konieczności stosowania zaawansowanego sprzętu komputerowego. Zwykle wystarczą procesory lub mikrokontrolery taktowane niezbyt szybkimi zegarami [3].

### 1. ROZWÓJ SYSTEMÓW TSR

Systemy TSR tworzy się z wykorzystaniem metod wizualnego przetwarzania obrazów (ang. Methods in Computer Vision). Architektura tych systemów może zależeć od ich przeznaczenia. Wg [3] w każdym systemie TSR możemy wyróżnić moduł detekcji i moduł klasyfikacji. Zadaniem detektora może być informowanie modułu klasyfikacji, jaki rodzaj znaku został wykryty (np. czy jest to trójkątny znak ostrzegawczy, czy okrągły znak zakazu). Detektor może reagować zarówno na kształt znaku, jak i na jego kolorystykę. Najważniejszym modułem systemu TSR jest moduł klasyfikacji, ponieważ od tego modułu zależy zazwyczaj niezawodność systemu. Klasyfikator implementowany jest przy użyciu sieci neuronowych, których wybór zależy od specyfiki zadania. Może on wykrywać jeden rodzaj obiektu (np. systemy wykrywające ograniczenia prędkości na drodze, reagujące na jeden rodzaj znaku) lub kilka rodzajów obiektów (systemy potrafiące wykryć znaki drogowe o różnych kształtach). Oprócz sieci neuronowych w klasyfikacji obrazów stosuje się deskryptory obrazów. Najpopularniejsze deskryptory to HOG (ang. Histogram of Oriented Gradients), Generalized Haar Wavelets, PCA-SIFT, Shape Contexts. W niniejszej pracy użyty został deskryptor HOG, który stanowi reprezentację cyfrowego obrazu w postaci wektora o  $n$  współrzędnych. W deskrytorze zachowana jest informacja o kształcie przedmiotów znajdujących się w obrazie. Długość  $n$  wektora będącego wynikiem wyznaczania HOG zależy od sposobu obliczania gradientu, wymiarów komórki, liczby bloków, jak również ich kształtu i sposobu nakładania się. Czynniki te zostały dokładniej opisane w podrozdziale 2. Deskryptor HOG może zostać przekazany bezpośrednio do sieci neuronowej SVM w celu jej nauczania lub testowania. W systemie badanym w niniejszym artykule można wyróżnić jedynie moduł klasyfikacji. W celach badawczych etap detekcji został pominięty. W dalszej części artykułu terminem detekcja sygnalizatora świetlnego będzie określany fakt

wykrycia przez system sygnalizatora świetlnego. W niektórych systemach TSR trudno jest odróżnić etap detekcji od etapu klasyfikacji [3].

## 2. DESKRYPTORY HOG

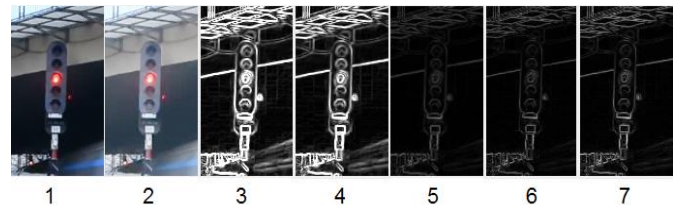
Deskryptor HOG (*ang. HOG Feature Vector*) powstaje po przekodowaniu cyfrowego obrazu na wektor o wymiarach  $1 \times n$ , gdzie wartość  $n$  zależy od sposobu kodowania deskryptora. Proces kodowania składa się z następujących etapów: normalizacja gamma obrazu z kamery, obliczenie obrazu gradientowego w celu wyodrębnienia konturów przedmiotów znajdujących się w obrazie po normalizacji gamma, podział obrazu gradientowego na komórki składające się z wektorów gradientu odpowiadających pikselom w obrazie przed przetworzeniem, kwantyzacja wektorów gradientu w komórkach, pogrupowanie komórek po kwantyzacji w nakładające się bloki, normalizacja bloków oraz przekształcenie bloków w wektor. Przekodowanie pełnego obrazu z kamery np. o rozdzielczości 640x480 pikseli dałoby w rezultacie długi wektor HOG. W celu oszczędności pamięci zazwyczaj po obrazie uzyskanym z kamery przesuwane jest okno o wymiarach 128x64 [1] lub w systemie badanym w niniejszym artykule – o wymiarach 160x96 pikseli, dla którego obliczany jest wektor HOG. Jest ono nazywane oknem detektora (*ang. detector window*).

Normalizację gamma obrazu wyjściowego przeprowadza się w celu wyeliminowania efektów oświetlenia (*ang. illumination effects*). W niniejszym artykule wykorzystywane były obrazy RGB (*ang. Red Green Blue*) o trzech kanałach kolorystycznych: czerwonym, zielonym i niebieskim. Każdy z kanałów przechowywany jest w postaci macierzy liczbowej o wymiarach odpowiadających liczbie pikseli obrazu w wierszach i kolumnach. Każda liczba w takiej macierzy odpowiada poziomowi nasycenia koloru w skali od 0 do 255 odpowiadającego danemu kanałowi kolorystycznemu. Normalizacja gamma polega na podzieleniu każdej liczby w macierzy odpowiadającej poszczególnemu kanałowi kolorystycznemu przez pierwiastek z sumy kwadratów liczb odpowiadających poszczególnemu pikselowi obrazu w macierzach z każdego kanału kolorystycznego. Operacja normalizacji gamma została przedstawiona we wzorze (3.1).

$$P_{norm} = \frac{P}{\sqrt{P_R^2 + P_G^2 + P_B^2}} \quad (3.1)$$

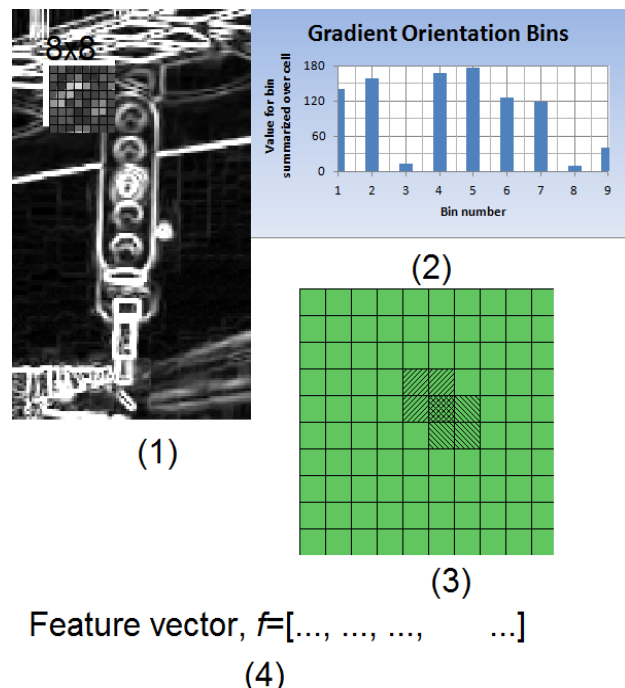
gdzie:  $P$  – liczbową wartość poziomu nasycenia koloru z normalizowanego kanału kolorystycznego dla wybranego piksela,  $P_R$ ,  $P_G$ ,  $P_B$  – liczbowe wartości poziomu nasycenia kolorów z kanałów kolorystycznych: czerwonego, zielonego i niebieskiego odpowiednio dla wybranego piksela,  $P_{norm}$  – znormalizowana wartość koloru dla wybranego piksela. Dla przyspieszenia obliczeń, w mianowniku wzoru (3.1) można umieścić sumę poziomów nasycenia kolorów dla trzech kanałów kolorystycznych.

Wyznaczanie obrazu gradientowego może być przeprowadzone na różne sposoby. Podczas wyznaczania gradientu można zastosować różne rodzaje operatorów np. Sobel'a, Prewitt'a, operator różnic centralnych, operator różnic średnich lub operator Roberta. Obraz wyjściowy, obraz po normalizacji gamma oraz obrazy zawierające długości wektorów gradientu (*ang. gradient magnitudes*) dla różnych operatorów gradientowych przedstawione zostały na rysunku 1.



Rys. 1. Okno detektora: obraz wyjściowy (1), po normalizacji gamma (2), gradientowy dla operatora Sobel'a (3), Prewitt'a (4), różnic centralnych (5), różnic średnich (6), Roberts'a (7)

Po wyznaczeniu obrazu gradientowego, obraz dzielony jest na komórki. Dla każdej komórki jest następnie wyznaczany wektor zawierający skwantowane wartości gradientów. Wektor zawiera  $k$  współrzędnych odpowiadających poziomom kwantyzacji (*ang. orientation bins*). Poziomy kwantyzacji gradientu mogą zmieniać się w zakresie od 0 do 180° lub od 0 do 360°. Od ich liczby zależy efektywność detektora. Ostatnim etapem jest normalizacja blokowa. Efektywność detektora zależy od liczby komórek w bloku oraz od ilości nakładających się komórek. W im większym stopniu bloki będą się nakładać, tym więcej danych nadmiarowych będzie zawierać finalny wektor HOG i tym więcej pamięci będzie on zajmował. Dla danego bloku wektory kwantyzacji z każdej komórki są łączone w jeden wektor, który następnie jest normalizowany. Normalizacja powoduje, że wektor HOG nie jest wrażliwy na zmiany oświetlenia w obrazie (*ang. illumination effects*). Etapy wyznaczania deskryptora HOG po uzyskaniu obrazu gradientowego przedstawione zostały na rysunku 2.



Rys. 2. Kolejne etapy wyznaczania deskryptora HOG. Podział obrazu gradientowego na komórki (1), kwantyzacja wektorów gradientów z komórki oraz przekształcenie ich w wektor (2), grupowanie komórek w nakładające się bloki (3), grupowanie wektorów kwantyzacji z bloków w finalny deskryptor HOG (4)

## 3. SIECI SVM

W niniejszym artykule przedstawione zostały badania detektora kolejowych sygnalizatorów świetlnych. Zaimplementowany przez autora detektor służy do wykrywania obiektów jednego rodzaju.

W trakcie detekcji przemieszczane jest okno detektora o rozmiarze 160x96 w obrębie klatki obrazu. Dla każdego położenia tego okna analizowany jest znajdujący się w jego obrębie fragment obrazu. W tym celu obliczany jest deskryptor HOG tego fragmentu a następnie na jego podstawie podejmowana jest decyzja czy zawiera on sygnalizator czy też jest tłem bez sygnalizatora. Decyzję tę podejmuje wytrenowany klasyfikator. Dla binarnej klasyfikacji bardzo dobrze nadają się sieci neuronowe SVM, które zostały wykorzystane do klasyfikacji obiektów w niniejszym artykule. Detektor sygnalizatorów świetlnych opisany w niniejszym artykule zawiera pojedynczy, liniowy klasyfikator SVM. Zbiór uczący zawierał 8000 przykładów fotografii tła o rozmiarze odpowiadającym oknu detektora uzyskanych z filmu video nagranych z kamery umieszczonej w lokomotywie kolejowej oraz z 8000 fotografii o tym samym rozmiarze zawierających sygnalizatory świetlne występujące na polskich szlakach kolejowych. W zbiorze testowym wykorzystano trzy obrazy o wymiarach 467x915, 512x919 oraz 470x922 pikseli. W procesie testowania detektora, po sekwencyjnym przesuwaniu okna detektora z odstępem 8 pikseli w poziomie oraz po każdym wierszu z odstępem 8 pikseli w pionie, po trzech obrazach testowych powstaje 12708 wektorów HOG, które przekazywane są kolejno do klasyfikatora SVM. Badany w niniejszym artykule detektor kolejowych sygnalizatorów świetlnych zaimplementowany został w środowisku Matlab. Do uczenia klasyfikatora użyta została funkcja *fitcsvm()*. W celu klasyfikacji deskryptorów HOG pochodzących z okna detektora utworzono funkcję (4.1) pozwalającą na zmianę przesunięcia (ang. *bias*) płaszczyzny separującej zbiór deskryptorów kolejowych sygnalizatorów świetlnych oraz zbiór deskryptorów tła.

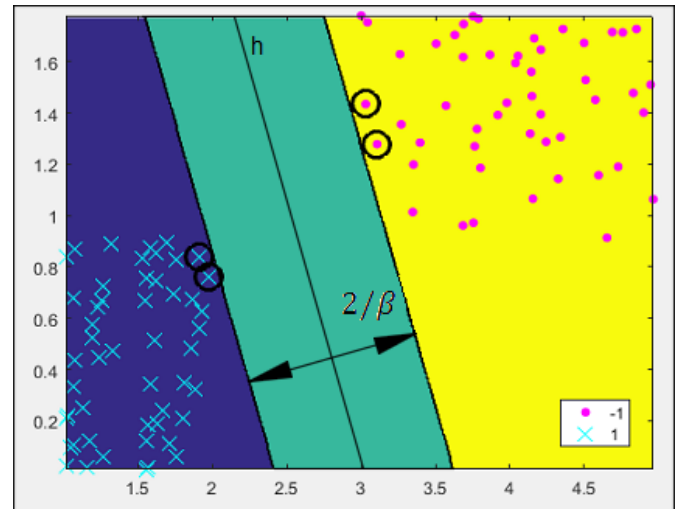
$$y = x^T \cdot \beta + b \quad (4.1)$$

gdzie:  $x$  – wektor deskryptora HOG utworzony z obrazu pochodzącego z okna detektora,  $\beta$  – wektor dopasowanych liniowych współczynników (utworzony w procesie uczenia klasyfikatora),  $b$  – przesunięcie hiperpłaszczyzny decyzyjnej, optymalna wartość tego parametru wyznaczana jest w procesie uczenia klasyfikatora,  $y$  – odległość wektora  $x$  od hiperpłaszczyzny decyzyjnej, która wyznaczana jest w procesie testowania; wartość  $y$  o znaku dodatnim informuje o sklasyfikowaniu wektora  $x$  w zbiorze sygnalizatorów świetlnych, wartość  $y$  o znaku ujemnym informuje o sklasyfikowaniu wektora  $x$  w zbiorze tła.

Możliwość zmiany parametru  $b$  pozwala na wyznaczenie krzywych kompromisu DET (ang. *Detection-Error Tradeoff*), które omówione zostały w rozdziale 4. Zasada działania liniowego klasyfikatora SVM na przykładzie punktów z przestrzeni  $R^2$ , gdzie hiperpłaszczyzna decyzyjna jest linią prostą rozdzielającą dwa zbiory, pokazana została na rysunku 3.

Punkty pokazane na rysunku 3 należą do zbioru uczącego klasyfikator. W niniejszym artykule odpowiadają one deskryptorom HOG dla zbioru sygnalizatorów oraz dla zbioru tła. Dodając kolejny punkt i sprawdzając jego położenie względem prostej  $h$  możemy zdecydować, do którego ze zbiorów powinien on należeć. Odpowiada to procesowi detekcji nowego punktu. W przypadku SVM oprócz parametru  $\beta$  oraz  $b$  określanych w procesie uczenia istotny jest również parametr  $C$ . Pozwala on na „ukrycie” danych uczących znajdujących się po niewłaściwych stronach hiperpłaszczyzny (ang. *outliers*), co ma miejsce gdy dane należą do dwóch grup liniowo nieseparowalnych [4]. Parametr  $C$  dobierany jest przez użytkownika indywidualnie dla konkretnego problemu klasyfikacyjnego. W badaniach przeprowadzonych w niniejszym artykule przyjęto wartość  $C=1$ . Dane wg [1] uzyskane zostały dla stałej  $C=0.01$ . Po przeprowadzeniu eksperymentów dla takiej samej wartości  $C=0.01$  nie

zauważono znaczących różnic w uzyskanych rezultatach względem eksperymentów przeprowadzonych przy  $C=1$ .



**Rys. 3.** Prosta decyzyjna  $h$ , margines separacji danych odpowiadający  $2/\beta$ , punkty należące do zbioru „-1” – oznaczone różowymi kropkami oraz punkty należące do zbioru „1” oznaczone błękitnymi krzyżykami. Klasyfikatorem SVM jest prosta  $h$ , punkty leżące najbliżej prostej decyzyjnej (ang. *Support Vectors*) oznaczone zostały czarnymi okręgami

#### 4. DETEKcja SYGNALIZATORÓw ŚWIETLNYCH

Właściwości detektora można opisać za pomocą współczynników wskazanych w tabeli 1. To, że system TSR wykrywa wszystkie znaki nie oznacza, że jest to dobry system [3]. Jednym ze współczynników opisujących efektywność tych systemów jest współczynnik FP (ang. *false positives*). Jest on równy liczbie próbek (okien detektora) nie zawierających poszukiwanego obiektu, ale wykrytych jako właściwy obiekt. Wysoka wartość tego współczynnika źle świadczy o efektywności systemu. Oprócz współczynnika FP stosuje się także współczynniki  $FPPF=FP/f$  (ang. *False Positives Per Frame*), gdzie  $f$  – liczba analizowanych klatek obrazu,  $FPTP=FP/TP$  (ang. *False/True Positive Ratio*),  $WPA=WP/AP$  (ang. *Wrong Pixels Per Area*), gdzie  $WP$  – liczba błędnie sklasyfikowanych pikseli,  $AP$  – całkowita liczba pikseli poddanych klasyfikacji [3]. Dodatkowo można wyróżnić 4 współczynniki utworzone ze współczynników zawartych w tabeli 1: TPR (ang. *true positive rate*), FNR (ang. *false negative rate*), FPR (ang. *false positive rate*) oraz TNR (ang. *true negative rate*). Współczynniki te zostały zapisane w równaniach (5.1) – (5.4) odpowiednio.

**Tab. 1.** Współczynniki opisujące detektor

Rodzaj obrazu w oknie detektora (próbka)	Próbka wykryta jako właściwy obiekt	Próbka wykryta jako niewłaściwy obiekt
Obraz zawierający właściwy obiekt	True Positive (TP)	False Negative (FN)
Obraz nie zawierający właściwego obiektu	False Positive (FP)	True Negative (TN)

$$TPR = \frac{TP}{TP + FN} \quad (5.1)$$

$$FNR = \frac{FN}{TP + FN} \quad (5.2)$$

$$FPR = \frac{FP}{FP+TN} \quad (5.3)$$

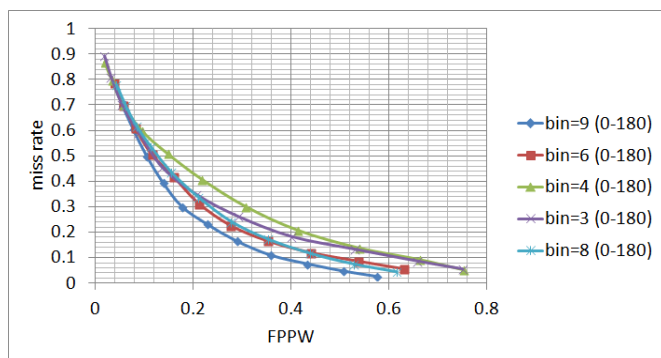
$$TNR = \frac{TN}{FP+TN} \quad (5.4)$$

Wartości współczynników zawartych w tabeli 1 odpowiadają liczbie zdarzeń, które zaszły w procesie detekcji, odpowiadających ich definicji. Współczynniki zawarte w równaniach (5.1)-(5.4) wyrażone są współczynnikami z tabeli 1 i są nazywane odpowiednio: współczynnik prawdopodobieństwa detekcji (*ang. recall*), współczynnik pominięcia właściwego obiektu lub współczynnik strat (*ang. miss rate*), prawdopodobieństwo fałszywego alarmu (*ang. fall-out*), współczynnik poprawnej klasyfikacji niewłaściwych obiektów (*ang. specificity*).

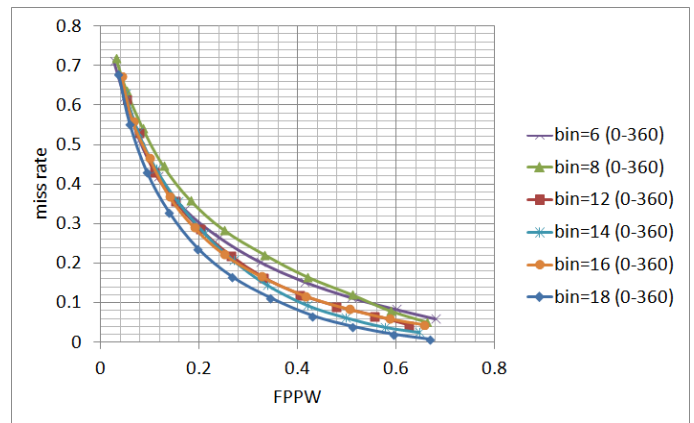
Badania przeprowadzone w niniejszym artykule wzorowane były na badaniach [1] przeprowadzonych dla detektora ludzi. Na podstawie uzyskanych rezultatów sporządzono krzywe DET, które wyrażają zależność współczynnika FNR (*ang. miss rate*) od współczynnika FPPW (*ang. false positives per window*) oznaczającego liczbę okien detektora zawierających tło, a wykrytych jako sygnalizator w stosunku do ilości okien przesłanych do klasyfikatora w całym procesie testowania klasyfikatora. W niektórych publikacjach zamiast współczynnika FPPW stosuje się współczynnik FPPI (*ang. false positives per image*) różniący się jedynie tym, że współczynnik FP jest obliczany dla pojedynczego obrazu testowego.

Badania przeprowadzone zostały dla domyślnych ustawień detektora zawartych w tabeli 2. Dla każdego wykresu zmieniano różne składowe deskryptora HOG: liczbę poziomów kwantyzacji, sposób nakładania się bloków oraz ich wymiary. Sporządzony został także wykres słupkowy dla kombinacji wymiarów komórki i wymiarów bloku, który przedstawia współczynnik strat (*miss rate*) przy współczynniku FP=0.15.

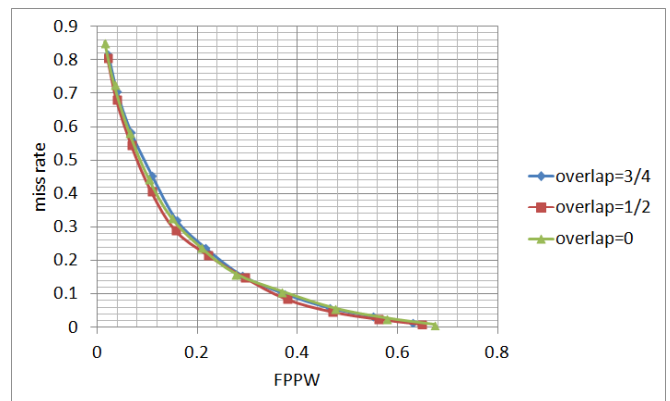
Na rysunkach 4-7 przedstawiono krzywe DET dla różnych parametrów detektora. Na rysunku 8 przedstawiono współczynnik *miss rate* dla różnych parametrów detektora przy FP=0.15. Rysunki 9 i 10 zawierają porównanie krzywych DET uzyskanych przez autora artykułu z badaniami Navneet Dalal [1] przeprowadzonymi dla detektora ludzi.



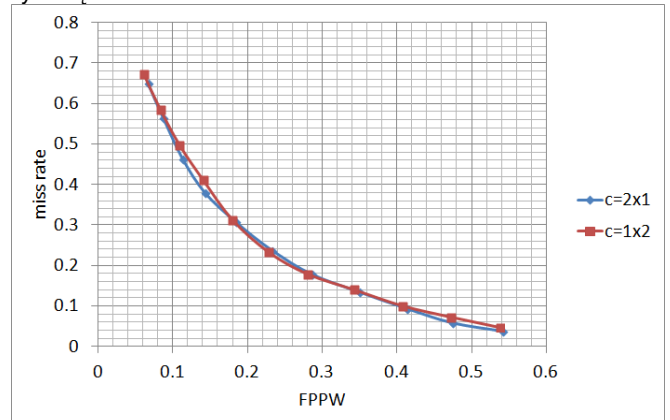
**Rys. 4.** Krzywe DET dla różnej liczby poziomów kwantyzacji przy orientacji poziomów kwantyzacji 0-180°



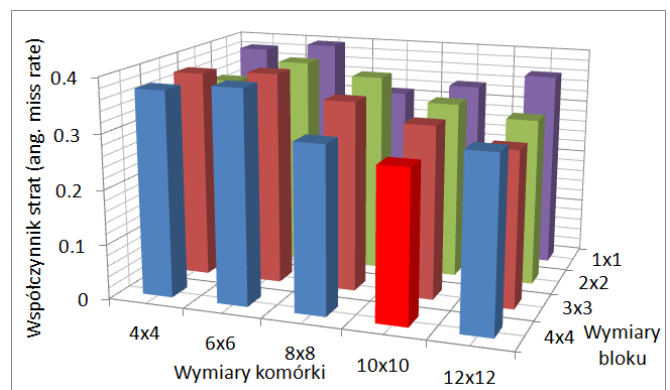
**Rys. 5.** Krzywe DET dla różnej liczby poziomów kwantyzacji przy orientacji poziomów kwantyzacji 0-360°



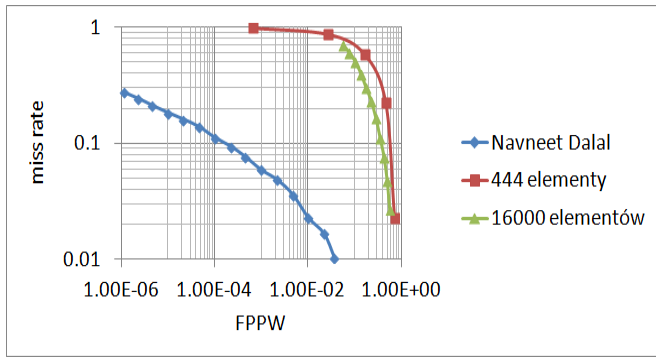
**Rys. 6.** Krzywe DET dla bloków komórek bez nałożenia, nakładających się w 1/2 oraz w 1/3



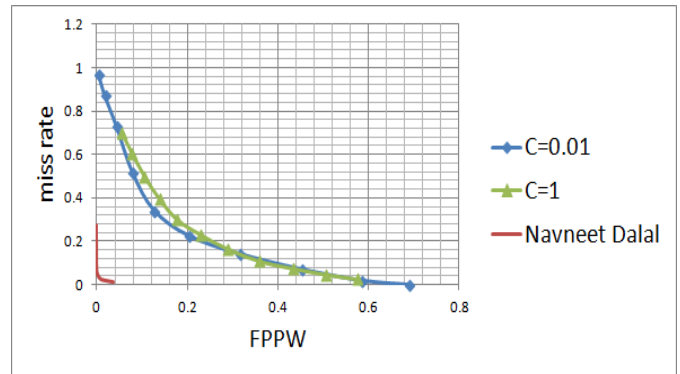
**Rys. 7.** Krzywe DET dla bloków pionowych o wymiarach 2x1 komórki oraz poziomych o wymiarach 1x2 komórki



**Rys. 8.** Wartość współczynnika *miss rate* przy współczynniku FP=0.15, dla różnych kombinacji wielkości komórki oraz bloku



Rys. 9. Porównanie krzywych DET dla domyślnych ustawień detektora wg publikacji Navneet Dalal [1], dla zbioru 222 sygnalizatorów świetlnych oraz dla zbioru 8000 sygnalizatorów świetlnych



Rys. 10. Krzywe DET dla domyślnych ustawień detektora (tabela 2) przy wartości marginesu separacji  $C=0.01$  oraz  $C=1$  dla badań autorów w porównaniu z badaniami Navneet Dalal [1]

Tab. 2. Domyślne ustawienia detektora

Rozmiar okna detektora	Liczba poziomów kwantyzacji wektorów gradientu w komórce	Orientacja poziomów kwantyzacji	Rozmiar komórki (w pikselach)	Rozmiar bloku (w komórkach)	Liczba komórek nakładających się sąsiednich w blokach (margines)
160x96	9	0-180°	8x8	2x2	1

**PODSUMOWANIE**

Na rysunkach 4 i 5 przedstawiony jest współczynnik miss rate w zależności od współczynnika FPPW, przy różnych poziomach kwantyzacji wektorów gradientu. Przy orientacji poziomów kwantyzacji 0-180° dla 3 poziomów kwantyzacji, przy FPPW=0.15 uzyskano o 6% lepszy współczynnik miss rate niż dla 4. Dla 8 poziomów kwantyzacji uzyskano podobne rezultaty jak dla 6. Korzystne jest stosowanie 3, 6 i 9 poziomów kwantyzacji, natomiast wartości poziomów kwantyzacji będące wielokrotnością liczby 2 wypadają słabo. Podobną zależność można zaobserwować na rysunku 5, gdzie przedstawione są takie same rezultaty, ale dla orientacji poziomów kwantyzacji 0-360°. Dla 6 poziomów kwantyzacji uzyskano wartość współczynnika miss rate o 4% niższą (przy FPPW=0.15) niż dla 8 poziomów. Krzywa dla 12 poziomów kwantyzacji pokrywa się z krzywą dla 16 poziomów kwantyzacji. Przy FPPW większym od 0.32 współczynnik miss rate jest dla 14 poziomów kwantyzacji o 2% niższy niż dla 16. Wg badań Navneet Dalal [1] zwiększanie liczby poziomów kwantyzacji dla orientacji 0-180° zmniejszało współczynnik miss rate, co świadczy o wzroście efektywności detektora wraz ze zwiększaniem się liczby poziomów kwantyzacji. W publikacji [1] autor podkreśla także, że zmiana orientacji poziomów kwantyzacji z 0-180° na 0-360° pogarsza efektywność. W niniejszej pracy podczas badań procesu detekcji sygnalizatorów świetlnych uzyskano dla 6 poziomów kwantyzacji (przy orientacji 0-180°) wartość miss rate 0.42 dla FPPW=0.15, natomiast dla 6 poziomów kwantyzacji (przy orientacji 0-360°) wartość 0.37 dla FPPW=0.15. Przy FPPW=0.4 dla 6 poziomów kwantyzacji uzyskano wartość współczynnika miss rate 1.4 dla 0-180° oraz 1.6 dla 0-360°. W badaniach detektora ludzi przeprowadzonych przez Navneet Dalal [1] większa liczba poziomów kwantyzacji gwarantowała lepszą efektywność detektora. W badaniach sygnalizatorów świetlnych przeprowadzonych w niniejszym artykule takiej zależności nie ma. Dobrą efektywność można uzyskać jedynie dla wybranej liczby poziomów kwantyzacji.

W badaniach Navneet Dalal [1] najniższy współczynnik miss rate (10.4%), przy FPPW=10<sup>-4</sup> osiągnięto dla wymiarów komórki 6x6 pikseli oraz wymiarów bloku 3x3 komórki. Navneet Dalal w [1] zauważa interesujący fakt, że w stosowanych przez niego obrazach w oknie detektora o wymiarach 128x64, ludzkie kończyny mają zazwyczaj szerokość 6 pikseli oraz długość 18 pikseli. W niniejszej

pracy podczas badań sygnalizatorów świetlnych najniższy współczynnik miss rate (27.7%) przy FPPW=0.15 uzyskano dla wymiarów komórki 10x10 pikseli i wymiarów bloku 4x4 komórki.

W przedstawionych rezultatach badań na uwagę zasługuje wysoka wartość współczynnika FPPW. Przyczyną mogą być rozbieżności w sposobie wyznaczania deskryptora HOG w publikacji [1] i w niniejszej pracy, gdzie był on wyznaczany przy użyciu funkcji bibliotecznej ze środowiska Matlab – *extractHOGFeatures()*. W pracy [1] przedstawiono wykres krzywej DET o wysokich wartościach FPPW. Na wykresie porównywane były różne rodzaje normalizacji bloków deskryptora HOG. Krzywa o wysokich wartościach FPPW odpowiadała w tej pracy deskryptorowi HOG bez normalizacji blokowej.

Brak normalizacji mogą potwierdzać rezultaty przedstawione na rysunkach 6 i 7. Na rysunku 6 przedstawiono krzywe DET dla różnej liczby komórek w obszarach nakładania się bloków. Najniższe wartości współczynnika miss rate powinny być dla największej nadmiarowości danych w deskrytorze HOG, tzn. dla bloków nakładających się w ¾. W uzyskanych rezultatach, nieznacznie lepszą efektywność detektora uzyskano dla bloków nakładających się w ½ w stosunku do bloków nie nakładających się, a najslabszą efektywność dla bloków nakładających się w ¼.

Zaskakujący jest także brak wyraźnej różnicy w wartościach współczynnika miss rate dla krzywych DET przedstawionych na rysunku 7. Sygnalizatory świetlne mają (w przypadkach badanych w niniejszym artykule) orientację pionową, więc pionowa orientacja bloków 2x1 komórek powinna dać lepsze rezultaty niż orientacja pozioma (1x2 komórki).

Przyczyn dużych wartości współczynnika FPPW próbowano dopatrywać się innej wartości parametru  $C = 1$ , przy którym zostały wyznaczone dane w niniejszym artykule. Dane w publikacji [1] zostały uzyskane dla  $C=0.01$ . W niniejszym artykule przeprowadzono badania dla domyślnego detektora (tabela 2) przy  $C=0.01$  i nie otrzymano znaczących różnic w stosunku do danych obliczonych przy  $C=1$  (rysunek 10).

Znaczącą poprawę efektywności detektora uzyskano zwiększając liczbę próbek w zbiorze uczącym klasyfikator: z 444 do 16000. Rezultaty uzyskane w niniejszym artykule wyznaczone zostały dla 8000 próbek sygnalizatorów świetlnych oraz 8000 próbek tła w zbiorze uczącym. W publikacji [2] przedstawiającej te same badania, co w publikacji [1] użyto w zbiorze uczącym 2478 obrazów

ludzi, które powstały z 1239 obrazów poprzez ich symetryczne odbicia oraz 12180 obrazów tła.

## BIBLIOGRAFIA

1. N. Dalal, *Finding People in Images and Videos*, THESE pour obtenir le grade de Docteur de L'Institut National Polytechnique de Grenoble, 17 Juillet 2006
2. N. Dalal, B. Triggs, *Histograms of Oriented Gradients for Human Detection*, <http://lear.inrialpes.fr>
3. A. Mogelmoose, M. M. Trivedi, T. B. Moeslund, *Vision based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems*, IEEE Transactions on Intelligent Transportation Systems vol.13, No.4, December 2012, s.(1484-1497)
4. C.J.C Burges, *A tutorial on Support Vector Machines for pattern recognition*, Data Mining and Knowledge Discovery, vol.2, Kluwer Academic Publishers, 1998, s.(121-167)

### Computer vision based detection of railway traffic lights in images using HOG descriptors and SVM neural networks

*In July 2006 Navneet Dalal published the research of the detector constructed with linear SVM neural networks and HOG (Histogram of Oriented Gradients) descriptors which was used for detection of people in images and videos. In this paper similar research was done for detection of railway*

*traffic lights. TSR (Traffic Sign Recognition) system was implemented in Matlab and examined for various settings of HOG descriptors.*

*Despite the TSR system for railway traffic lights detection has the same settings as people detection system, the results are different. Good results was obtained only for selected number of bins in HOG descriptor, while in people detection the miss rate coefficient decreases with increasing number of bins.*

*Increasing number of pictures in learning set of SVM network from 444 to 16000 significantly improved effectiveness of the detector. Unfortunately, high value of FPPW (False Positives Per Window) coefficient was obtained in comparison to the people detection research. It can be caused by the lack of block normalization or different block normalization method in the Matlab function which computes HOG feature vector. High values of FPPW can also point that this kind of descriptor or methodology is improper for railway traffic lights detection.*

Autorzy:

dr inż. **Roman Król** – Uniwersytet Technologiczno-Humanistyczny im. Kazimierza Pułaskiego w Radomiu

dr hab. inż. **Piotr Bojarczak**, prof. UTH – Uniwersytet Technologiczno-Humanistyczny im. Kazimierza Pułaskiego w Radomiu