TECHNIKA TRANSPORTU SZYNOWEGO

Jacek JAKUBOWSKI, Jolanta PACAN

PRZETWARZANIE OBRAZU POLA OBSERWACJI METODĄ DOPASOWANIA DESKRYPTORÓW PUNKTÓW KLUCZOWYCH

Streszczenie

W artykule przedstawiono wyniki wstępnych badań nad możliwością wykorzystania transformacji obrazu SIFT w zagadnieniach interpretacji treści obrazu cyfrowego. Efektem transformacji jest zbiór punktów kluczowych, których opis wyrażony 128-elementowym wektorem cech stanowi dane wejściowe dla procedury klasyfikacji minimalno-odległościowej. Prezentowany materiał omawia własności samej metody oraz ilustruje w sposób ilościowy jej zdolność do detekcji wyróżnionej klasy obiektów, których wzorce znajdują się w bazie danych.

WSTĘP

Analiza obrazów wizyjnych stanowi podstawę funkcjonowania systemów widzenia maszynowego, których zadaniem jest rozpoznanie i interpretacja zarejestrowanego cyfrowo obrazu na poziomie porównywalnym z percepcją człowieka. Nietrudno wyobrazić sobie aplikacje, w których pożądane jest poddanie analizie obrazów przedstawiających obszary strzeżone celem automatycznego wykrycia niepożądanych obiektów w polu widzenia kamery i uruchomienia alarmu. Dotyczyć to może np. miejsc przed bankami, dworców, lotnisk itp., gdzie obecność pojazdów może rodzić określony stan zagrożenia. Pożądanym w tym zakresie jest również automatyczny monitoring cech charakterystycznych obiektów przemieszczających się w strefie chronionej umożliwiający przeprowadzanie czynności dochodzeniowych lub działania prewencyjne. Z punktu widzenia obróbki komputerowej obraz stanowi macierz dyskretnych i ograniczonych wartości, uporządkowanych w sposób ukazujący rozkład intensywności na płaszczyźnie detekcji matrycy światłoczułej. Przetwarzanie takiej macierzy obejmuje zazwyczaj tzw. przetwarzanie niskiego rzędu, czyli przetwarzanie wstępne, którego pierwszym celem jest korekcja układu rejestracji jak np. korekcja jasności czy filtracja. Realizacja szczegółowych zadań systemu widzenia maszynowego wymaga jednak zastosowania operacji przetwarzania wysokiego rzędu, zależnych i dostosowanych do zakładanego celu (detekcja, klasyfikacja i ocena własności stanu obiektu, śledzenie itp.). Jedną z najnowszych współczesnych metodologii postępowania w tym zakresie jest poszukiwanie punktów charakterystycznych, które mogłyby zostać użyte do oceny stopnia zgodności pomiędzy dwoma obrazami tej samej scenerii lub tego samego obiektu. Obszary jej zastosowań obejmują syntezę zdjęć panoramicznych i lotniczych, orientowanie przestrzenne robotów mobilnych i wyszukiwanie znanych wzorców w obrazach cyfrowych (np. wzorców obrazów twarzy). Artykuł przedstawia próbę oceny zdolności uogólniania takiego podejścia, tzn. możliwości realizacji zadania wykrywania obecności nie

konkretnego obiektu, ale wybranej kategorii obiektów na analizowanym obrazie pola obserwacji na przykładzie kategorii samochodów osobowych. Wykorzystana metoda bazowa nosi nazwę SIFT (ang. *Scale Invariant Feature Transform*) i obok podobnych metod jak np. SURF czy MSER powoli staje się standardem w obszarze przetwarzania obrazów [2][4][5]. Do badań użyta została dostępna w Internecie baza danych Uniwersytetu Illinois o nazwie „*UIUC Image Database for Car Detection*" autorstwa [1].

1. APARAT MATEMATYCZNY

1.1. Wprowadzenie

W przedmiotowym procesie rozpoznawania na podstawie dyskretnych punktów, które można wykorzystać do oceny podobieństwa obiektów występujących na obrazach, wyróżnia się 3 etapy. Pierwszym etapem jest wykrycie na obrazie tzw. punktów kluczowych, do których zalicza się specyficzne konfiguracje pikseli układające się w pewne struktury jak np. punkty narożne, punkty, w których zlokalizowane są obszary o charakterze ciemnych plam na jasnym tle lub odwrotnie (tzw. *blobs*), punkty będące zakończeniami linii, punkty występowania obszarów o kształcie litery T itp. Najbardziej pożądaną własnością detektora punktów kluczowych jest jego powtarzalność, którą można określić jako zdolność do wykrywania tych samych punktów obiektu na obrazie przy zmieniających się warunkach akwizycji, do których zalicza się zmianę skali, obrót, przesunięcie, rozmycie, perspektywę, częściowe przesłonięcie obiektu, zmiany kontrastu. Powtarzalne wykrywanie punktów kluczowych jest determinowane własnościami samych punktów. Punkty wykrywane z dużym stopniem niezależności od transformacji lub zniekształcenia obiektu na obrazie określa się mianem punktów stabilnych. Etap drugi to wytworzenie wektora cech opisujących otoczenie punktów charakterystycznych. Deskryptor cech powinien być wysoce dystynktywny i podobnie jak detektor – odporny na deformacje geometryczne i fotometryczne obrazu. Etap ostatni polega na dopasowaniu (stwierdzeniu zgodności) wektorów cech różnych obrazów z wykorzystaniem pewnych miar odległości. Na czas przebiegu operacji dopasowania bezpośredni wpływ ma wymiar wektora cech. Wektory krótkie zapewniają uzyskiwanie dużych szybkości przetwarzania, ale są jednocześnie mniej dystynktywne.

1.2. Detekcja punktów kluczowych

Realizacja etapu pierwszego – detekcji punktów charakterystycznych odbywa się formalnie z wykorzystaniem filtracji obrazu. Stosowny filtr daje odpowiedzi o dużej wartości dla pikseli należących do detekowanych struktur i małe dla pikseli do nich nie należących. Cechą wspólną punktów kluczowych, wykorzystywaną w ich detekcji, jest występowanie dużych zmian luminancji obrazów. Metody detekcji zmian luminancji można zgrupować w dwie kategorie. Pierwsza grupa metod wykorzystuje badanie pierwszej pochodnej (metody gradientowe), gdzie detekcja odbywa się na drodze wyszukiwania jej ekstremów. W drugiej grupie badaniu podlega druga pochodna, czyli szybkość zmian gradientu a konkretnie jej przejście przez zera. Z uwagi na własność niezmienniczości względem obrotu, do najpopularniejszych detektorów w zadaniu wykrycia punktów charakterystycznych obrazów należą detektory wykorzystujące informację o drugiej pochodnej – detektory bazujące na Laplasjanie oraz macierzy Hessego. Metody gradientowe tej własności nie posiadają.

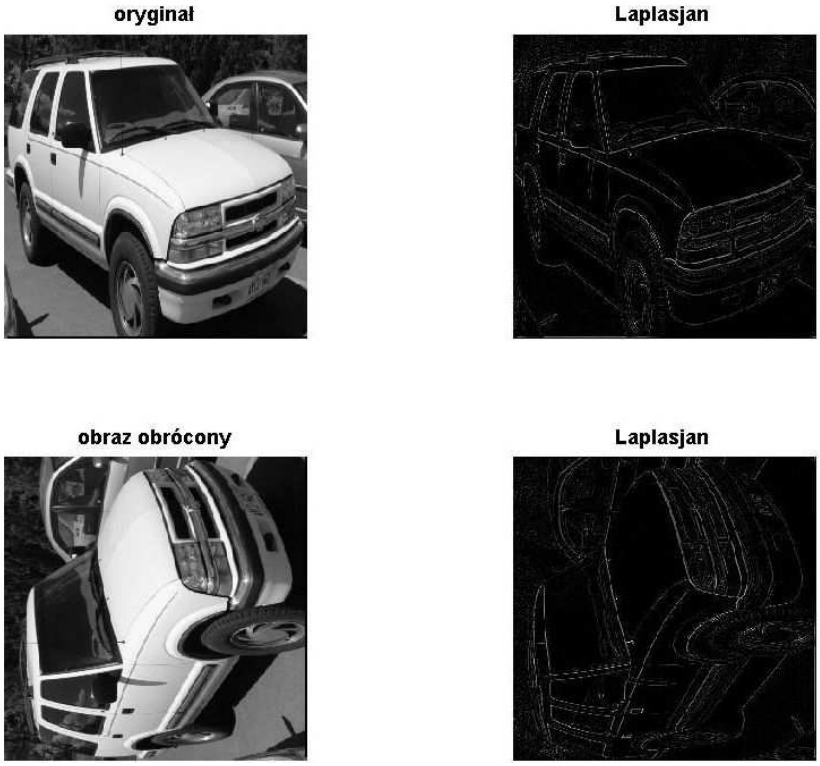
Wykorzystywany w metodzie SIFT Laplasjan (a dokładniej jego aproksymacja) jest odpowiednikiem pochodnej dwuwymiarowej i dla obrazu traktowanego właśnie jako obiekt dwuwymiarowy $I(\mathbf{x})=I(x,y)$ jest dla każdej lokalizacji piksela $\mathbf{x}= (x,y)^T$ sumą drugich pochodnych cząstkowych na kierunkach x i y:

$$\Delta I(x,y) = \frac{\partial^2 I(x,y)}{\partial x^2} + \frac{\partial^2 I(x,y)}{\partial y^2}. \tag{1}$$

Z uwagi na właściwości drugiej pochodnej, Laplasjan charakteryzuje się uzyskiwaniem wartości ekstremalnych w okolicach krawędzi obrazu. Dyskretny charakter obrazu nie pozwala na dokładne określenie Laplasjanu, ale możliwe jest zastosowanie przybliżenia różnicowego, co w praktyce wiąże się z filtracją obrazu wejściowego z filtrem (maską) o postaci:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{2}$$

Wynik zastosowania powyższego filtru do rzeczywistego obrazu oraz obrazu obróconego przedstawia rys. 1.



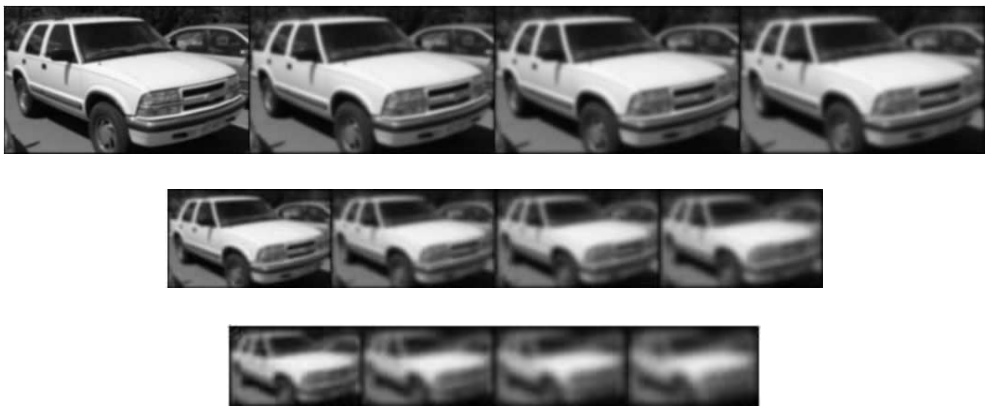
Rys.1. Detekcja krawędzi za pomocą Laplasjanu.

Widoczna jest pozytywna cecha Laplasjanu polegająca na detekcji krawędzi niezależnie od obrotu obrazu. Z powodu oczywistej wrażliwości drugiej pochodnej na obecność szumu, stosowaną praktycznie techniką jest poprzedzająca wyznaczenie Laplasjanu wstępna filtracja obrazu za pomocą filtru gaussowskiego, co prowadzi do odmiany Laplasjanu znanej pod nazwą Laplasjan filtru Gaussa – LoG (ang. *Laplacian of Gaussian*). Dodatkowymi parametrami wpływającymi na wynik przetwarzania jest tutaj rozmiar filtru (maski) oraz jego szerokość, będąca odpowiednikiem odchylenia standardowego rozkładu Gaussa. Oprócz naturalnej zdolności detektorów opartych na Laplasjanie do wykrywania krawędzi, dzięki filtracji Gaussa możliwe jest również ich użycie do detekcji obiektów o charakterze plam,

czyli *blob*-ów. W wykorzystanej na potrzeby niniejszej pracy metodzie SIFT, ze względu na możliwość przyspieszenia obliczeń stosowana jest aproksymacja filtru LoG za pomocą różnicy dwu filtrów Gaussa – DoG (ang. *Difference of Gaussian*) [6]. Wykorzystanie tej aproksymacji warunkowane jest wytworzeniem tzw. przestrzeni skal, czyli sekwencji obrazów powstałych na drodze filtracji badanego obrazu $I(x,y)$ przez filtry Gaussa $G(x,y,\sigma)$ o różnych wartościach odchylenia standardowego σ. W wyniku filtracji, powstaje obraz $L(x, y, \sigma)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)\,, \tag{3}$$

którego stopień rozmycia zależy od wartości parametru σ. Im większe odchylenie standardowe filtru gaussowskiego, tym bardziej rozmyty obraz. Natomiast im bardziej rozmyty obraz, tym mniej szczegółów jest na nim widocznych – rys. 2. Zatem parametr σ może być utożsamiany z odległością obiektu od rejestratora obrazu a tym samym ze skalą.



Rys. 2. Przestrzenie skal zgrupowane w 3 oktawy.

Pozyskanie punktów kluczowych charakteryzujących się odpornością na szeroki zakres zmian skali możliwe jest dzięki filtracji kaskadowej, która prowadzi do powstania tzw. oktaw przestrzeni skal – rys. 2. Kolejne oktawy tworzone są przez powtórzenie procedury kaskadowej filtracji po przeprowadzeniu decymacji ostatniego obrazu gaussowskiego z oktawy poprzedniej. Obrazy w każdej kolejnej oktawie mają dzięki temu dwukrotnie mniejszą rozdzielczość. Podstawą detekcji punktów kluczowych w metodzie SIFT jest wykorzystywanie obrazów będących różnicami dwóch sąsiednich obrazów $L(x,y,\sigma)$, powstałych przez filtrację funkcjami Gaussa. Na drodze porównania wartości danego piksela ze wszystkim sąsiadującymi punktami rozpatrywanego obrazu (8 pikseli) oraz obrazów znajdujących się powyżej i poniżej w przestrzeni skal (po 9 pikseli w każdym), w obrazach różnicowych wyszukiwane są ekstrema. Aby rozważany punkt uznać za ekstremum, jego wartość musi być mniejsza lub większa od wartości wszystkich 26 sąsiadujących punktów. Spośród punktów ekstremalnych, na drodze selekcji odrzucającej traktowane jako niestabilne punkty o małym kontraście i punkty leżące na krawędziach, dokonuje się wyboru punktów kluczowych.

1.3. Opis otoczenia punktów kluczowych

Niezależność opisu punktu kluczowego od obrotu uzyskuje się na drodze wyznaczenia tzw. orientacji, czyli dominującego kierunku lokalnego gradientu obrazu w otoczeniu danego punktu. Sam opis (deskryptor) otoczenia punktu kluczowego jest uzyskiwany na podstawie

modułów gradientów z najbliższego sąsiedztwa 16x16, których kierunki wyznaczane są względem określonej wcześniej orientacji. Sąsiedztwo to dzieli się na mniejsze obszary o wymiarach 4x4 każdy, gdzie dla 8 kierunków rozmieszczonych równomiernie w zakresie od 0 do 360 stopni wyznaczane są wypadkowe moduły gradientu. Wektor cech stanowi zestawienie modułów gradientów każdego z obszarów – łącznie 4x4x8, czyli zawiera 128 wartości. Szczegóły transformacji można znaleźć w [4].

1.4. Dopasowanie punktów kluczowych

Rozpoznawanie na podstawie deskryptorów punktów kluczowych zasadza się na przyjęciu postulatu, że są one cechami dystynktywnymi obiektów, które opisują i w praktyce polega na minimalno-odległościowym dopasowaniu punktów kluczowych z obrazu wejściowego do punktów w bazie wzorców [4]. Prostą i skuteczną metodą na takie dopasowanie jest metoda najbliższego sąsiedztwa, która w wariancie 1-NN wskaże dla danego punktu kluczowego dokładnie jeden punkt z bazy wzorców. Po przebadaniu wszystkich punktów kluczowych danego obrazu możliwe jest wyznaczenie dla niego histogramu ich przynależności i podjęcie decyzji o zaklasyfikowaniu obiektu. W przypadku, gdy w bazie wzorców znajdują się punkty kluczowe danego obiektu i obiekt ten jest już wykryty i wycięty z szerszego pola obserwacji za pomocą innych metod, powyższa metodyka postępowania zapewnia uzyskanie zadowalających wyników rozpoznawania [3]. Podejmowane w niniejszym referacie próby dotyczą zadania trudniejszego gdyż związanego z faktem, że w bazie wzorców znajdują się punkty kluczowe nie tych samych obiektów, ale obiektów jedynie podobnych, które tworzą tym samym pewną wyróżnioną kategorię.

2. MATERIAŁ

Wykorzystana w pracy baza danych UIUC zawiera dwie kategorie danych. Pierwszą, którą można traktować jako kategorię danych treningowych stanowi zbiór 1050 obrazów o jednakowych rozmiarach obejmujący 550 przypadków widoku bocznego różnych marek aut osobowych oraz 500 przypadków widoków nie-aut. Zbiór drugi zawiera 165 obrazów widoków bocznych aut ale o większym w stosunku do obrazów treningowych udziale tła.

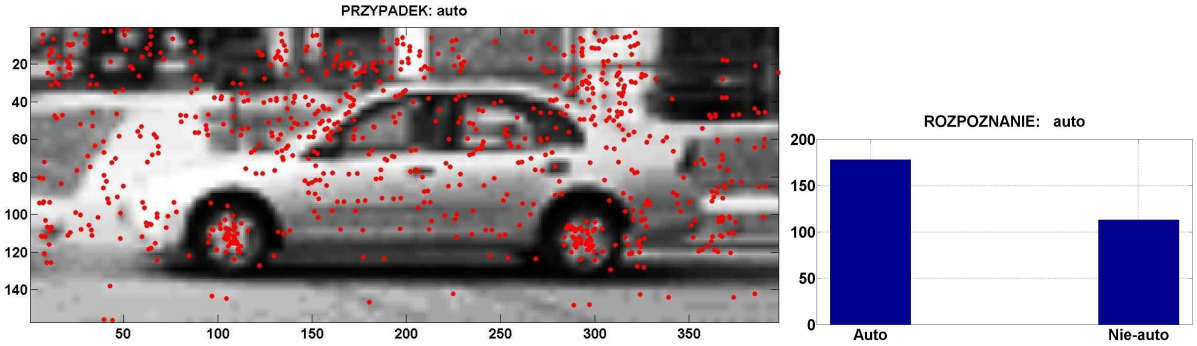
3. METODYKA BADAŃ I WYNIKI

Podstawą detekcji wyróżnionej klasy obiektów jest umieszczenie w bazie danych deskryptorów punktów kluczowych znanych obiektów wzorcowych z którymi w drugim kroku porównywane są punkty kluczowe obrazu testowego. Ze względu na praktyczne aspekty podejmowania decyzji o wykryciu obiektu w różnych warunkach akwizycji, badania skuteczności takiej operacji przeprowadzono dla dwu następujących przypadków, które można utożsamiać z rosnącym stopniem jej trudności:

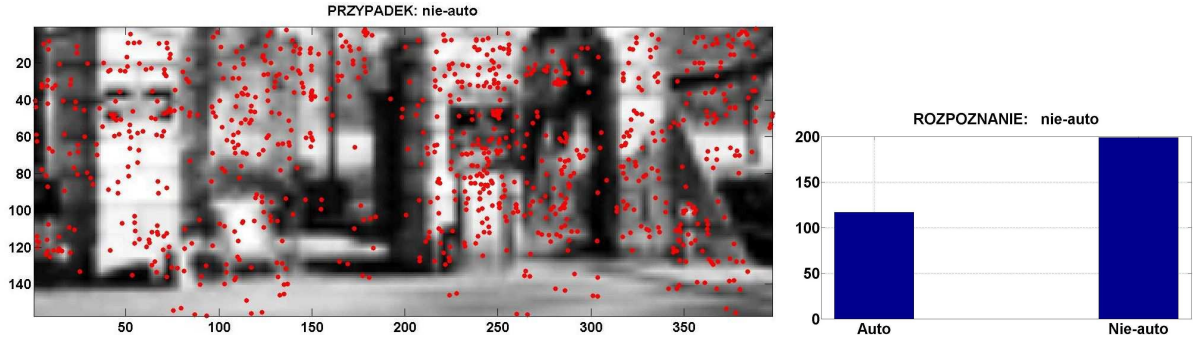
1. Obrazy testowe mają takie same rozmiary jak obrazy uczące, na podstawie których tworzona była baza danych i jednocześnie wykrywane obiekty zawarte na obrazach testowych mają zbliżoną skalę do obiektów z obrazów uczących.
2. Obrazy testowe różnią się od obrazów uczących zarówno rozmiarami jak i skalą wykrywanych obiektów.

Podjęcie decyzji o stwierdzeniu bądź nie obecności obiektu dla przypadku 1 postuluje się rozwiązać na podstawie charakteru histogramu przynależności punktów kluczowych danego obrazu do jednej z dwu klas, których przykładowe przypadki znajdują się w bazie. Histogram taki można utworzyć korzystając ze wspomnianej metody 1-NN (jednego najbliższego sąsiada) poprzez przypisanie danemu punktowi kluczowemu tylko jednej etykiety klasy z bazy danych, a mianowicie etykiety tej klasy, która zawiera punkt położony najbliżej. W

przypadku, gdy wzorce badanego nowego przypadku znajdują się w bazie, to należy się spodziewać koncentracji przynależności jego punktów kluczowych do jednej klasy. Procedurę wykrycia można wówczas zrealizować na podstawie prostego głosowania: jeśli liczba punktów kluczowych przypisanych do klasy pierwszej jest większa od liczby punktów kluczowych przypisanych do klasy drugiej, to wynikiem rozpoznania jest klasa 1 i na odwrót. Badania dla przypadku 1 przeprowadzone zostały metodą kroswalidacyjną z wykorzystaniem zbioru pierwszego bazy danych UIUC, którego 80% obrazów zostało użytych do wytworzenia danych uczących a 20% do testowania. W proponowanym klasyfikatorze nie istnieje faza uczenia – klasyfikacja przebiega na bieżąco na drodze wyszukiwania tych zgromadzonych w bazie przypadków, które odpowiadają przypadkom nowym. W efekcie, celem oceny możliwości prezentowanej metody, błędy rozpoznawania wyznaczane były wyłącznie na podstawie tych losowo wybranych obrazów testowych, które nie wchodziły do bazy wzorców. Przykładowe obrazy testujące z każdej z dwu klas (klasa *auto* i klasa *nie-auto*) wraz przypisanymi do nich histogramami przynależności ich punktów kluczowych do wektorów z bazy danych przedstawiają rysunki 3 i 4.



Rys. 3. Rozkład punktów kluczowych na obrazie i histogram przynależności do klas dla przypadku: auto. Zwraca uwagę duża koncentracja punktów kluczowych w okolicach kół pojazdu.



Rys. 4. Rozkład punktów kluczowych na obrazie i histogram przynależności do klas dla przypadku: nie-auto.

Należy zwrócić uwagę na istnienie w tym przypadku możliwości stosunkowo poprawnego podjęcia decyzji o wykryciu obiektu na podstawie porównania wysokości słupków w histogramach przynależności punktów kluczowych. Ilościowe zestawienie wyników wykrycia przedstawia Tabela 1.

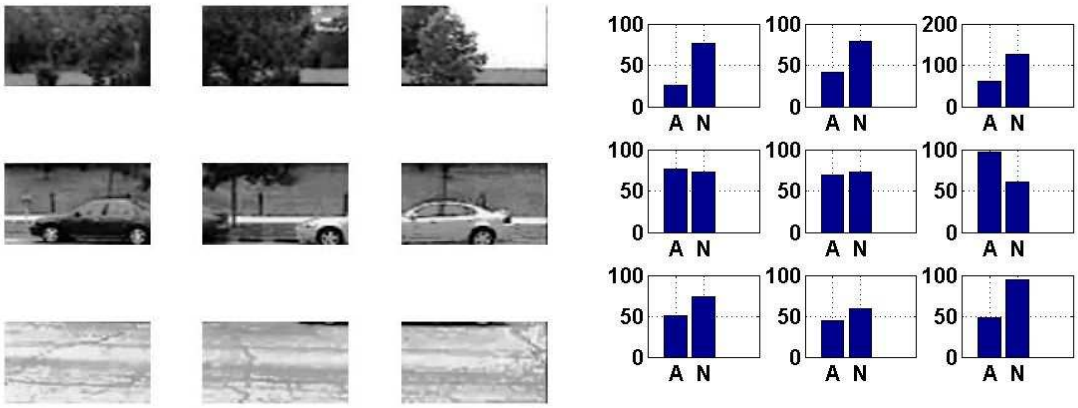
Rozwiązanie przedmiotowego zadania powyższą metodą dla przypadku 2 nie jest jednak możliwe. Zwiększone rozmiary obrazu badanego w stosunku do obrazów wzorcowych powodują, że liczba punktów kluczowych wykrytych w tle jest na ogół większa od liczby punktów kluczowych, które można przypisać w obrazie badanym do wykrywanego obiektu.

Wobec metodologii 1-NN, która zawsze znajdzie najbliższego sąsiada, wynik głosowania punktów kluczowych będzie w większości przypadków wskazywał na klasę nie-auto.

Tab. 1. Wyniki wykrycia dla przypadku 1

rodzaj obiektu	liczba przypadków	liczba błędnych rozpoznań	błąd względny
auto	300	16	5.3%
nie-auto	300	20	6.7%
sumarycznie	600	36	6%

Naturalne zmniejszenie liczby punktów kluczowych można uzyskać dla podobrazów powstałych na drodze podziału obrazu badanego na kilka mniejszych. Każdy z nich, oddzielnie od pozostałych, może być wówczas poddany operacji dopasowania do dostępnych w bazie wzorców punktów, dzięki czemu unika się nadmiaru punktów tła. Ideę wykrywania zakładanego obiektu dla przypadku wyodrębnienia siatki podobrazów 3x3 ilustruje rys. 5.



Rys. 5. Podział obrazu badanego na podobrazy z odpowiadającymi im histogramami przynależności do klas (oznaczenie: A – auto, N – nie auto).

Każdy z wyciętych podobrazów charakteryzuje się własnym histogramem przynależności do klas, który decyduje o wyniku detekcji. W przypadku, gdy słupek odpowiadający przynależności do klasy auto jest wyższy od słupka odpowiadającego klasie nie-auto, to wyciętemu podobrazowi można przypisać obecność wykrywanego obiektu. Wynik detekcji dla przypadku analizowanego na rys. 5 przedstawiony jest na rys. 6.

Rys. 6. Wyniki detekcji aut w przykładowym obrazie testowym uzyskane z wykorzystaniem proponowanej metody i wydzieleniu siatki podobrazów o rozmiarach 3x3.

Krytyczny dla rozpatrywanej metody jest dobór siatki podziału na podobrazy. Zbyt duże podobrazy nie eliminują w sposób dostateczny punktów tła, natomiast zbyt małe stanowią źródło dużej ilości przypadkowych dopasowań. Celem statystycznej oceny proponowanej metody postępowania, przeprowadzone zostały badania wykonane z wykorzystaniem obrazów testowych bazy UIUC oraz własnych obrazów nie zawierających wykrywanej kategorii obiektów. W badaniach tych jako wskaźnik detekcji zastosowano sumę stosunków wysokości słupków histogramów tych podobrazów, w których metoda wykryła obecność wyróżnionych obiektów. Otrzymane wyniki wskazują na uzyskiwanie błędu detekcji na poziomie ok. 20%.

PODSUMOWANIE

Zaprezentowany materiał przedstawia wstępne wyniki badań nad zagadnieniem rozpoznawania scenerii zawartej w obrazie cyfrowym. Wykorzystana metoda dopasowania deskryptorów punktów kluczowych badanego obrazu do punktów kluczowych bazy wzorców dowodzi istnienia potencjalnej możliwości wykrycia wyróżnionej klasy obiektów. Przedstawione wyniki wskazują jednak na konieczność dalszych prac nad zmniejszeniem uzyskanego stosunkowo dużego poziomu błędów detekcji. Wydaje się, że powinny one objąć wykorzystanie nie tylko informacji o wyniku porównania liczby dopasowanych punktów kluczowych, ale również informację o ich odległości od wzorców.

IMAGE SCENERY PROCESSING WITH THE USE OF MATCHING BASED ON KEY POINT DESCRIPTORS

Abstract

The paper presents the results of an initial research on the possibilities to use SIFT transform as a method to analyze the scene in digital images. The output of the transform is a set of key points described with a 128-element vector of features that can be used as an input to a minimum distance classifier. Presented material shows basic properties of the method as well as its quantitative assessment to detect distinguished objects of known patterns included in the data base.

BIBLIOGRAFIA

1. Agarwal S., Awan A., Roth D.: *Learning to Detect Objects in Images via a Sparse, Part-Based Representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, Nr. 11, 2004, ss. 1475-1490.
2. Bay H., Ess A., Tuytelaars T., Van Gool L.: *Speeded-Up Robust Features (SURF)*, Computer Vision and Image Understanding, 110(2008), ss. 346-359.
3. Jakubowski J.: *Ocena możliwości wykorzystania deskryptorów cech lokalnych obrazu twarzy w zadaniu automatycznej identyfikacji osób*, Przegląd Elektrotechniczny, R. 88, NR 9a/2012, ss. 217-221.
4. Lowe D.: *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision, Nr **60** (2), ss. 91–110,
5. Matas J, Chum O., Urban M., Pajdla T.: *Robust wide baseline stereo from maximally stable extremal regions*, Proc. of British Machine Vision Conference, ss. 384-396, 2002.

6. Mikolajczyk K., Schmid C., *Scale & Affine Invariant Interest Point Detectors*, International Journal of Computer Vision, vol. 60 (1), 2004, ss. 63-86.

Autorzy:
dr inż. Jacek JAKUBOWSKI – Wojskowa Akademia Techniczna
mgr inż. Jolanta PACAN – Wojskowa Akademia Techniczna