

WYKORZYSTANIE TRANSFORMACJI LOG-HOUGH'A DO TWORZENIA REPREZENTACJI OBRAZU DLA KLASYFIKATORA NEURONOWEGO

LOG-HOUGH BASED IMAGE REPRESENTATION FOR THE NEURAL CLASSIFIER

Przemysław Piekarski, Zbigniew Mikrut

Katedra Automatyki, Akademia Górniczo-Hutnicza w Krakowie

SŁOWA KLUCZOWE: fotogrametria, reprezentacja obrazu, log-polar, log-Hough, wzajemne dopasowanie, sieci *backpropagation*

STRESZCZENIE: Artykuł przedstawia metodę tworzenia reprezentacji fragmentu obrazu opartą o transformacje log-polar i log-Hough'a. Transformacje te są uważane za uproszczone modele biologicznych systemów wizyjnych. Reprezentację obrazu stanowią rzuty przestrzeni log-Hough'a na osie: kątów i promieni. Tak utworzone wektory stanowią wejście do sieci neuronowej typu *backpropagation*. Zadaniem sieci jest klasyfikacja reprezentacji obrazów na „korzystne” i „niekorzystne” z punktu widzenia późniejszego dopasowywania, którego celem jest automatyczna orientacja wzajemna zdjęć fotogrametrycznych. Badano sieci z jedną warstwą ukrytą o zmiennej liczbie elementów. Najlepsze sieci rozpoznały zbiór testowy na poziomie 70%.

1. WSTĘP

Jednym z podstawowych etapów przygotowania modelu fotogrametrycznego jest tzw. orientacja wzajemna modelu, czyli wyznaczenie wzajemnego położenia kolejnych zdjęć lotniczych. Procedura ta składa się z wielu etapów (Heipke, 1997), zawsze jednak pierwszym krokiem jest wybór odpowiednio informatywnych obszarów, na których następnie wyszukiwane są punkty homologiczne. Zwykle takie obszary wyznacza się manualnie.

Celem badań była próba automatyzacji tego etapu. W tym celu należało dokonać wyboru odpowiedniej reprezentacji fragmentu obrazu a następnie sprawdzić, czy wybór ten był słuszny. Sprawdzenia zwykle dokonuje się w sposób praktyczny: wektory reprezentujące obrazy podaje się na wejście wybranych klasyfikatorów. Jeśli rozpoznania plasują się na wysokim poziomie to uznaje się, że badana reprezentacja została wybrana prawidłowo. W prezentowanej pracy do testów wybrano jeden z najmocniejszych i najczęściej używanych klasyfikatorów – sieć neuronową typu *backpropagation*.

Do utworzenia reprezentacji wybrano dwie mniej znane transformaty – log-polar i log-Hough'a, które są przybliżonym modelem niższych pięter systemu wzrokowego ssaków (Schwartz, 1977; Weiman, 1989). Głównymi zaletami transformacji log-polar są: duża redukcja danych, przekształcanie rotacji i powiększania obiektu w przestrzeni

kartezjańskiej na równoległe przesunięcia wzdłuż odpowiednich osi w przestrzeni log-polar, oraz duża rozdzielczość obrazu w centrum przy równoczesnym dużym kącie widzenia. Zalety te umożliwiają dokonywanie precyzyjnej normalizacji obrazu, zarówno pod względem położenia kąтового jak i względem powiększenia. Praktyczne zastosowania transformaty log-polar można znaleźć w pracach (Lee *et al.*, 1993; Sandini *et al.*, 1992; Wilson *et al.*, 1992; Tadeusiewicz *et al.*, 1998; Kubiak *et al.*, 2004). Matematyczny opis transformaty log-polar zamieszczono w rozdziale 2.

Transformata log-Hough'a wykrywa odcinki linii prostych i łagodne łuki na krawędziowym obrazie log-polar (Weiman, 1989). W przestrzeni log-Hough'a powstają maksima w miejscach odpowiadających parametrom wykrytych linii. Podobnie jak w transformacie Hough'a maksima te są tym wyższe im więcej współliniowych punktów znajdowało się na krawędziach w przestrzeni log-polar. Operacja ta jest podobna do przekształcenia realizowanego na jednej z warstw kory wzrokowej - wykrywania fragmentów linii o różnych nachyleniach.

Obie te transformaty zostały wybrane do skonstruowania reprezentacji obrazu ponieważ dokonują one ekstrakcji krawędzi oraz dostarczają w formie skondensowanej informacji o wzajemnym ich położeniu.

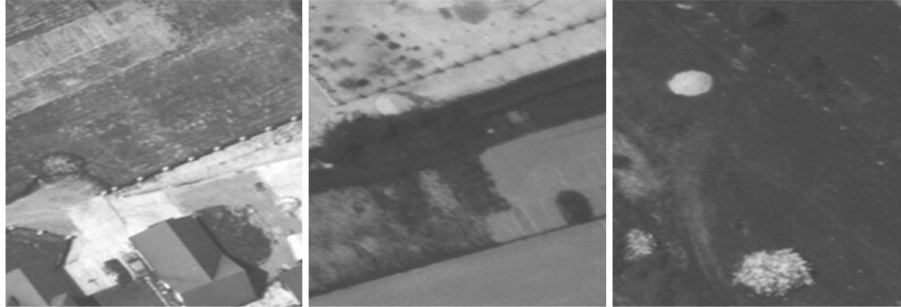
2. TRANSFORMACJA OBRAZU DO PRZESTRZENI LOG-HOUGH'A

Na fragmentach zdjęć lotniczych zdecydowano się poszukiwać odcinków linii prostych oraz łagodnych łuków - ze względu na częstość ich wykorzystywania w procesie korelacji wzajemnej zdjęć. Ze zdjęć lotniczych wybrano 452 obrazy o rozmiarach 80x80 pikseli. Następnie grupa specjalistów-fotogrametrów dokonała klasyfikacji wybranych obszarów. Efektem tej operacji było zaszeregowanie każdego z obrazów do jednej z trzech klas: obrazów korzystnych, pośrednich lub niekorzystnych z punktu widzenia późniejszego procesu dopasowywania. Ponieważ opinie ekspertów czasem różniły się (szczególnie w odniesieniu do klasy „pośredniej”) przyjęto, że jako klasyfikacja finalna zostanie przyjęta mediana z uzyskanych odpowiedzi. Przykłady zaklasyfikowanych obszarów przedstawiono na rysunku 1.

Klasę „korzystną” stanowiły obszary zawierające obiekty sztuczne, wyraźnie kontrastujące z otoczeniem oraz posiadające dobrze widoczne punkty załamania granic, gdyż są one chętnie wykorzystywane podczas pomiarów fotogrametrycznych. W tej klasie znalazło się 142 obrazy.

Klasa „niekorzystna” to obszary bez obiektów sztucznych: zwykle są to tereny odkryte, użytkowane rolniczo, jednolite, nie posiadające punktów jednoznacznie orientujących te obszary w terenie, o słabo widocznych granicach oraz obrazy niewyraźne. Do tej klasy zaliczono 221 obiektów.

Klasę „pośrednią” stanowiły obszary nie zaklasyfikowane do żadnej z w/w klas. Są to głównie obiekty rolne lub leśne, posiadające niewyraźne granice pomiędzy uprawami, obszary o granicach wyraźnych, lecz nie dających możliwości jednoznacznej lokalizacji w terenie. Ta klasa zawiera 89 obiektów.



Rys. 1. Przykłady obszarów (w kolejności od lewej): „korzystnego”, „pośredniego” i „niekorzystnego”

Ogólnie na algorytm przekształcania obrazu z przestrzeni kartezjańskiej do postaci reprezentacji w przestrzeni log-Hough'a składają się następujące etapy:

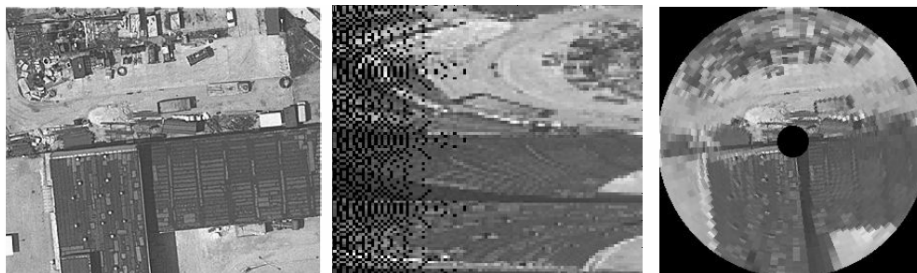
1. transformacja do przestrzeni log-polar (na tym etapie należało podjąć decyzję jakie parametry transformaty powinny być użyte),
2. wykrywanie krawędzi i binaryzacja (należało zbadać, który z licznych detektorów krawędzi zastosować),
3. przekształcenie do przestrzeni log-Hough'a,
4. zrzutowanie maksimum na osie, utworzenie wektorów wejściowych dla sieci neuronowej (na tym etapie należało skonstruować i oprogramować algorytmy normalizacji i agregacji wektorów).

Przykład przekształcenia obrazu fotogrametrycznego do przestrzeni log-polar przedstawiono na rysunku 2.

Transformację cyfrowego obrazu z układu kartezjańskiego (x, y) do przestrzeni log-polar (u, v) opisuje układ równań:

$$(1) \begin{cases} u = \ln(r) = \ln(\sqrt{x^2 + y^2}), r \geq R_0 > \varphi \\ v = \arctan\left(\frac{y}{x}\right) \end{cases}$$

przy czym r jest promieniem kolejnych współśrodkowych okręgów, a R_0 jest tzw. strefą martwą przekształcenia. Jeżeli promienie r będą rosły wykładniczo to po zastosowaniu do kolejnych pól receptorowych równań (1) otrzymuje się równomierne rozmieszczenie pikseli w przestrzeni log-polar. Ze względu na początkowy charakter badań zdecydowano się na taki dobór współczynników przekształcenia log-polar aby uzyskać obraz po transformacji o rozmiarach 50x50 pikseli, co oznacza, że na obraz kartezjański została nałożona siatka składająca się z 50 pierścieni podzielonych równomiernie na 50 fragmentów (kątown).



Rys. 2. Przekształcenie log-polar. W kolejności od lewej: obraz w przestrzeni kartezyjskiej – w przestrzeni log-polar (obraz kilkakrotnie powiększony) – po przekształceniu odwrotnym (z log-polar do układu kartezyjskiego)

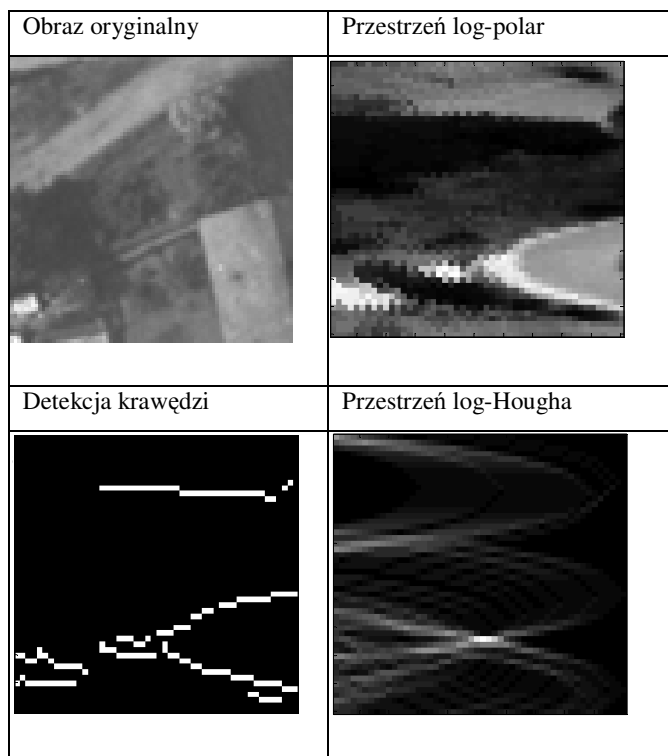
Transformata log-Hough'a jest odpowiednikiem transformaty Hough'a (Hough, 1962), realizowanej dla krawędziowego obrazu obiektu w przestrzeni log-polar. Akumulacyjna przestrzeń log-Hough'a jest geometrycznie identyczna z przestrzenią log-polar (Weiman, 1989), dzięki czemu implementacja transformaty jest szybka i dokładna.

Oprócz wymienionych wyżej kroków algorytmu (1-4) należy – w różnej kolejności i na różnych etapach algorytmu – realizować operacje przetwarzania cyfrowych obrazów, takie jak detekcja krawędzi czy też binaryzacja. Przetestowano trzy kombinacje, które przedstawiono w tabeli 1. Ze względu na niekorzystny wpływ operacji binaryzacji w przestrzeni kartezyjskiej na liczbę szczegółów (Metoda I), oraz niekorzystny wpływ redukcji rozdzielczości transformacji log-polar w stosunku do wykrytych na zdjęciu krawędzi (Metoda II), zdecydowano się na etapy przygotowania reprezentacji proponowane w metodzie III.

Tab 1. Charakterystyka badanych metod generacji reprezentacji obrazu

Nazwa metody	Przestrzeń kartezyjska		Przestrzeń log-polar	
	Binaryzacja	Detekcja krawędzi	Binaryzacja	Detekcja krawędzi
Metoda I	Tak	Nie	Tak	Nie
Metoda II	Nie	Tak	Nie	Tak
Metoda III	Nie	Nie	Nie	Tak

Przyjęty do realizacji ciąg operacji obejmuje kolejno: przekształcenie do przestrzeni log-polar, detekcję krawędzi, binaryzację i transformację log-Hough'a. Etapy te zostały zilustrowane na rysunku 3.



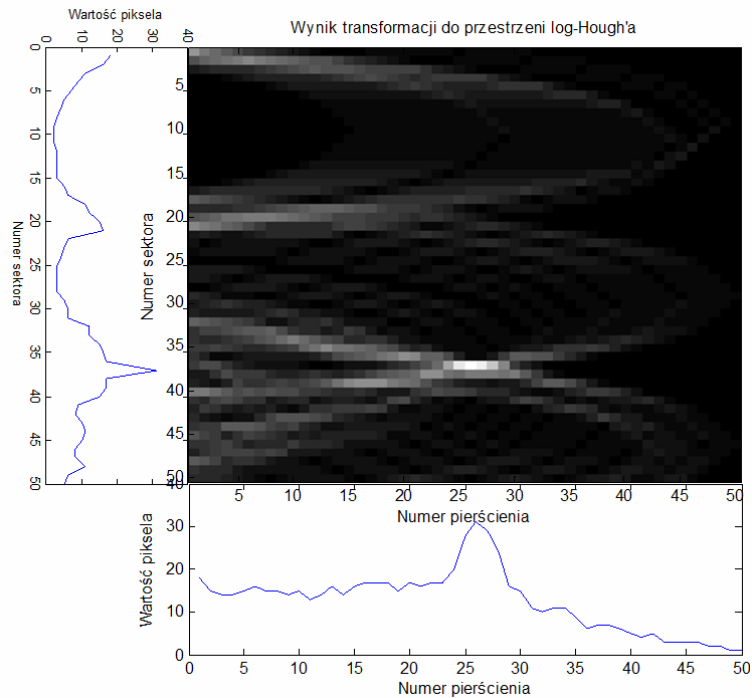
Rys. 3. Kolejność przekształceń obrazu – od obrazu kartezjańskiego do przestrzeni log-Hough'a

Obraz w przestrzeni log-Hough'a zawiera maksima, których wielkość w przybliżeniu odpowiada istniejącym w przestrzeni kartezjańskiej obiektom liniowym i łagodnym łukom. Końcowym etapem przygotowywania wektora cech dla sieci neuronowej było zrzutowanie wartości maksymalnych na oś pionową i poziomą (por. rys. 4). Pozwoliło to uzyskać dwa wektory, które – dla danego obrazu - charakteryzują zarówno długość jak i położenie krawędzi, a tym samym ich wzajemne usytuowanie.

We wstępnej fazie eksperymentów zbadano wpływ następujących czynników na postać otrzymanej reprezentacji:

- operatora detekcji krawędzi w przestrzeni log-polar (testowano operatory: Canny'ego, Laplacian of Gaussians, Prewitta, Robertsa, Sobela, Zero-cross), z których do dalszych badań wybrano metodę Prewitta,
- wyboru części lub całości reprezentacji (wektor odpowiadający osi kątów i/lub osi promieni), po wstępnych testach wybrano oś promieni (pierścieni),

- agregacji elementów wektora – w związku z dużą dokładnością odwzorowania i wyraźnie zarysowanymi maksimumami wektor zmniejszono dwukrotnie wybierając z dwóch sąsiednich elementów większy.



Rys. 4. Utworzenie dwóch wektorów poprzez rzutowanie maksimumów przestrzeni log-Hough'a na osie układu współrzędnych

3. KLASYFIKACJA PODOBRAZÓW PRZY POMOCY SIECI NEURONOWEJ

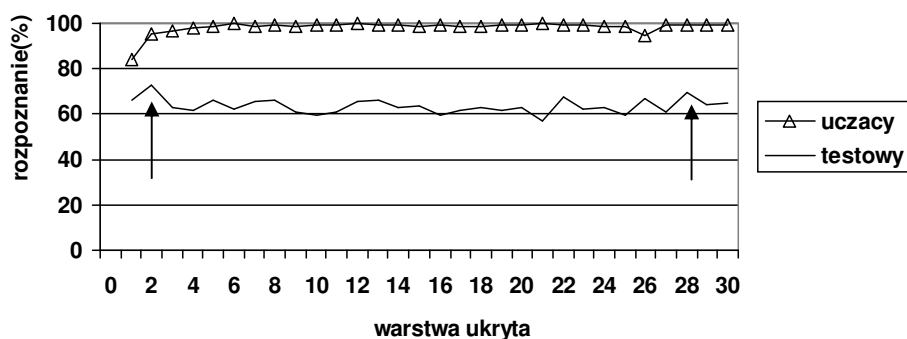
Utworzone w wyżej opisany sposób reprezentacje wektorowe stanowiły dane wejściowe do sieci neuronowej typu *backpropagation*, zaimplementowanej za pomocą oprogramowania PRTools (Duin, 2000; Matlab 2006), z którego wykorzystano opcje uczenia i rozpoznawania badanych obiektów siecią neuronową z jedną warstwą ukrytą o liczbie elementów zmieniającej się od 2 do 30. Z uwagi na brak pozytywnych rezultatów rozpoznawania trzech klas warstwa wyjściowa sieci składała się z dwóch neuronów odpowiadających dwóm rodzajom obrazów: „korzystnym” i „niekorzystnym”.

Przeprowadzono eksperymenty z dwoma zbiorami danych.

W pierwszym przypadku reprezentacje „pośrednie” obrazów dołączono do klasy „korzystne”. Dla rodziny sieci o warstwach ukrytych zmieniających się od 2 do 30 najlepsze osiągnięte wyniki to rozpoznanie zbioru testowego na poziomie 65% - dla sieci

o strukturach 25-20-2¹ oraz 25-29-2. Zbiór uczący został rozpoznany przez te sieci odpowiednio na poziomach 75% i 98%.

W drugim wariantcie reprezentacje klasy „pośredniej” całkowicie usunięto ze zbioru danych. Wyniki uczenia i rozpoznawania nieco się poprawiły (por. rys. 5). Dla dwóch sieci (o strukturach: 25-2-2 oraz 25-28-2) osiągnięto rozpoznanie zbioru testowego na poziomie 70% (zbiory uczące były rozpoznawane na poziomie 95% i 99%).



Rys. 5. Sieć *backpropagation* o zmiennej liczbie neuronów w warstwie ukrytej: wyniki uczenia i testowania drugim wariantem zbioru danych

Dla obu sieci uzyskano bardzo zbliżone macierze błędów (*confusion matrices*). Dla zbioru testowego miały one postać odpowiednio:

$$(2) \begin{bmatrix} 47 & 24 \\ 26 & 85 \end{bmatrix} \quad \begin{bmatrix} 45 & 26 \\ 30 & 81 \end{bmatrix} \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

przy czym elementy a_{11} obu macierzy (2) dotyczą rozpoznań elementów „korzystnych”, elementy a_{22} – obrazów „niekorzystnych”. Liczby położone na przeciwprzekątnej mają zbliżone (i dosyć duże) wartości. Nie jest więc możliwe dokonanie wyboru pomiędzy omawianymi sieciami.

4. WNIOSKI

W wyniku przeprowadzonych eksperymentów sprawdzających użyteczność reprezentacji obrazu opartej na transformatach log-polar i log-Hough'a otrzymano rezultaty poprawnej klasyfikacji podobrazów zdjęć lotniczych na poziomie 70%. Wynik ten może jeszcze być o kilka procent poprawiony jeśli weźmie się pod uwagę pewność rozpoznania i zastosuje tzw. odrzucenia części klasyfikowanych danych. Należy jednak stwierdzić, że

¹ notacja: liczba neuronów w warstwie wejściowej – ukrytej - wyjściowej

nawet poprawiony w ten sposób rezultat – choć pozytywny - nie jest satysfakcjonujący z punktu widzenia zastosowań praktycznych.

Biorąc pod uwagę doniesienia literaturowe raportujące pozytywne próby stosowania tych transformacji do tworzenia reprezentacji należy – zdaniem autorów – kontynuować badania, koncentrując się na zastosowaniu prezentowanej metody do mniejszych obszarów obrazu lub wręcz do wydzielonych uprzednio obiektów. Pozwoli to na uzyskanie bardziej wyrazistych (indywidualnych) reprezentacji, które być może staną się pomocne na kolejnym etapie analizy zdjęć lotniczych – etapie dopasowywania, czyli wyznaczania wzajemnej orientacji zdjęć fotogrametrycznych.

5. LITERATURA

Duin R.P.W., 2000. *PRTools Version 3.0 - a Matlab Toolbox for Pattern Recognition*. Delft University of Technology, <http://www.ph.tn.tudelft.nl/prtools>

Heipke C., 1997. *Automation of interior, relative and absolute orientation*. ISPRS Journal of Photogrammetry & Remote Sensing, 52, 1-19

Hough P.V.C., 1962. *Method and Means for Recognizing Complex Patterns*. US Patent 3069654.

Kubiak A., Mikrut Z., 2004. *Zastosowanie transformaty log-Hough'a i sieci neuronowych do identyfikacji połamanych ziaren pszenicy*. Automatyka AGH, t.8 z.3, 2004, str. 477-486

Matlab 2006. *Image Processing Toolbox For Use with Matlab. User's Guide*. The MathWorks, Inc., Natick.

Lee S., Choi Y., 1993. *Unconstrained handwritten zip code recognition*. Proc. WCNN, Portland.

Sandini G., Tistarelli M., 1992. *Vision and Space-Variant Sensing*. In: H. Wechsler ed., *Neural Networks for Perception*, Academic Press Inc., Boston.

Schwartz E.L., 1977. *Spatial mapping in the primate sensory projection: analytic structure and relevance to perception*. Biological Cybernetics, nr 25, s. 181-194.

Tadeusiewicz R., Mikrut Z., 1998. *Neural-Based Object Recognition Support: from Classical Preprocessing to Space-Variant Sensing*. Proc. of the ICSC/IFAC Symposium on Neural Computation '98, Vienna, Austria, 23-25 September, s. 463-468.

Weiman C.F.R., 1989. *Polar exponential sensor arrays unify iconic and Hough space representation*. SPIE vol.1192: Intelligent robots and computer vision VIII, 1989.

Wilson J.C., Hodgson, R.M., 1992. *Log-polar mapping applied to pattern representation and recognition*. Academic Press Inc.,

Wykonano w ramach grantu KBN nr 4T12E 001 27

LOG-HOUGH BASED IMAGE REPRESENTATION FOR THE NEURAL CLASSIFIER

KEY WORDS: photogrammetry, image representation, log-polar, log-Hough, mutual matching, *backpropagation* network

SUMMARY: In the present paper, the method for generation of the sub-image representation is presented. The method is based on log-polar and log-Hough transforms. These transforms are considered to be very simplified models of the biological visual systems. The projections of the log-Hough space onto the two axes (the angles and the radii ones) are taken as the sub-image representation. These vectors form an input to the *backpropagation* neural network. The network task is to classify the sub-images as "advantageous" or "non-advantageous" from the subsequent mutual matching point of view. Several networks which have a variable number of neurons in one hidden layer have been tested. The best recognition rates about 70% (test set) have been obtained.

mgr inż. Przemysław Piekarski
e-mail: przemek_pp@orange.pl

dr inż. Zbigniew Mikrut
e-mail: zibi@agh.edu.pl
telefon:+ 012 617 38 53