

Zbigniew Mikrut*

Tworzenie reprezentacji obszarów zdjęć lotniczych za pomocą sieci neuronowych klasy PCNN**

1. Wprowadzenie

Pomimo ciągłego wzrostu mocy obliczeniowej komputerów nadal istnieją obszary badawcze, w których metoda „brutalnej siły” nie do końca się sprawdza. Jednym z takich obszarów jest fotogrametria, a właściwie ta jej część, która zajmuje się analizą i wzajemnym dopasowywaniem sekwencji zdjęć lotniczych.

Lecący na stałej wysokości samolot wykonuje serie zdjęć, które pokrywają się w około 60%. Przyjmuje się, że skala fotografowanego terenu nie ulega zmianie, natomiast w praktyce możliwe są niewielkie odchylenia katowe kolejnych fotografii. Zadaniem osób (programów) dopasowujących kolejne zdjęcia jest wyznaczenie określonej liczby par odpowiadających sobie punktów. W tym celu – po zeskanowaniu fotografii – wybiera się obszary, zawierające informatywne punkty, wskazuje się je na jednym ze zdjęć a następnie odszukuje na drugim zdjęciu – najpierw cały obszar a następnie odpowiedniki punktów, wskazanych na zdjęciu pierwszym. Za pomocą zmian powiększenia obrazu można osiągnąć subpikselową dokładność parowania [8].

Obecnie problem znajdowania odpowiedników na obrazach lotniczych rozwiązuje się dwiema metodami. Pierwsza polega na stosowaniu wspomnianej wyżej „brutalnej siły”. Za pomocą odpowiedniego oprogramowania generowana jest na obrazie pierwszym prostokątna siatka. Następnie tak powstałe obszary są dopasowywane do fragmentów na drugim obrazie – zwykle metodą korelacyjną, która nie zawsze jest skuteczna. Drugie podejście to dopasowywanie „ręczne”: człowiek wybiera informatywne obszary, w nich punkty i wskazuje je na drugim obrazie.

Celem tej pracy jest dokonanie próby automatyzacji pierwszego etapu analizy, czyli zbadania, czy jest możliwe stwierdzenie, że wybrany fragment obrazu jest wystarczająco „informatywny” i nadaje się do dalszego przetwarzania. Można przyjąć, że jest to problem rozpoznawania, który sprowadza się do dwóch – w dużej mierze niezależnych od siebie – zadań: zdefiniowania odpowiedniej reprezentacji fragmentu obrazu i zastosowania efek-

* Katedra Automatyki, Akademia Górniczo-Hutnicza w Krakowie

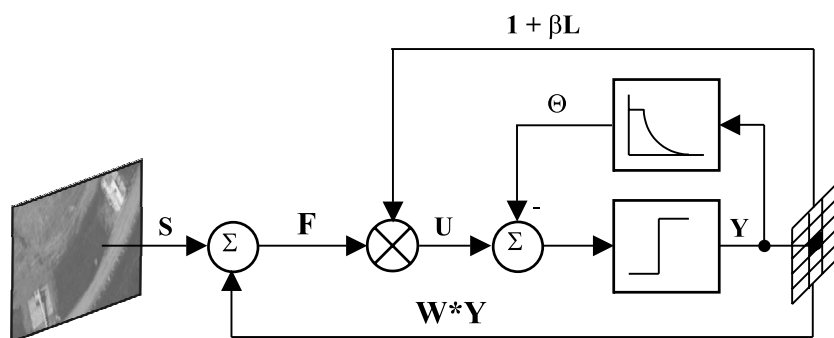
** Wykonano w ramach grantu KBN nr 4T12E 001 27

tywnego klasyfikatora. Zestawienie wyników prac różnych autorów i prezentacje konkretnych par propozycji rozwiązań „reprezentacja – klasyfikator” można znaleźć w pracy [10].

W ramach prezentowanych badań zdecydowano się na wykorzystanie stosunkowo nowej i mało znanej (lecz obiecującej) metody „sygnatur” (*image signatures*) jako reprezentacji fragmentu obrazu. Sygnatury są generowane przez impulsującą sieć neuronową. Zasada działania takiej sieci i metoda uzyskiwania sygnatur zostanie przedstawiona w rozdziale 2. Klasyfikatorem dla tak utworzonych reprezentacji będzie sieć neuronowa typu *backpropagation*, uważana za jedną z efektywniejszych i najczęściej używanych sieci [11].

2. Sieci impulsujące i sygnatury

Model sieci impulsujących PCNN (*Pulse Coupled Neural Networks*) jest wynikiem badań nad systemem wzrokowym kota, prowadzonych pod kierunkiem Eckhorna [3]. Sieć zaproponowana przez Eckhorna składa się z warstwy neuronów połączonych ze sobą dwoma rodzajami sprzężeń zwrotnych: *feeding* (F) i *linking* (L). Sposób połączeń schematycznie przedstawiono na rysunku 1. Rozmiar sieci jest równy wielkości przetwarzanego obrazu cyfrowego. Kombinacja wartości piksela obrazu S z sygnałami F i L tworzy sumaryczny sygnał U (*internal activity*), który jest porównywany z progiem neuronu Θ . Po przekroczeniu progu następuje generacja impulsu (wyjście neuronu Y przyjmuje wartość 1). W tym momencie próg rośnie, a później stopniowo opada do wartości spoczynkowej, umożliwiając generowanie następnych impulsów. Sygnały F i L odzwierciedlają aktywność sąsiednich neuronów poprzez odpowiednio zdefiniowane macierze współczynników. Wartości sygnałów F i L – podobnie jak próg – także maleją w miarę upływu czasu.



Rys. 1. Schemat blokowy modelu neuronu, tworzącego sieć PCNN

Równania opisujące działanie pojedynczego impulsującego neuronu mają postać:

$$F_{ij}[n+1] = fF_{ij}[n] + S_{ij} + V_F \sum_{kl} W_{ijkl} Y_{kl}[n] \quad (1)$$

$$L_{ij}[n+1] = lL_{ij}[n] + V_L \sum_{kl} M_{ijkl} Y_{kl}[n] \quad (2)$$

$$U_{ij}[n+1] = F_{ij}[n+1](1 + \beta L_{ij}[n+1]) \quad (3)$$

$$\Theta_{ij}[n+1] = g\Theta_{ij}[n] + hY_{ij}[n+1] \quad (4)$$

$$Y_{ij}[n+1] = \begin{cases} 1 & \text{gd}y \ U_{ij}[n+1] > \Theta_{ij}[n] \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad (5)$$

gdzie:

U_{ij} – jest stanem wewnętrznym neuronu o współrzędnych ij ,

S_{ij} – jest pobudzeniem (pikselem obrazu wejściowego przeskalowanego do zakresu $[0, 1]$),

F_{ij}, L_{ij} – składowe uwzględniające sprzężenia zwrotne (tzw. *feeding* i *linking*),

Θ_{ij} – jest progami neuronu,

Y_{ij} – to stan zewnętrzny neuronu (1 – jest impuls, 0 – brak impulsu),

W_{ijkl} – to współczynniki sąsiedztwa neuronu ij o współrzędnych kl dla F ,

M_{ijkl} – to współczynniki sąsiedztwa neuronu ij o współrzędnych kl dla L ,

V_F, V_L – to współczynniki wzmocnienia dla bezpośrednich sprzężeń F i L ,

n – jest numerem iteracji ($n = 1, \dots, N$).

Współczynniki f, g i l określają stałe czasowe zmian sygnałów F, Θ i L ($f, g < 1, f > g$). Współczynnik β definiuje wpływ połączenia L na stan wewnętrzny neuronu U . Wyjściem sieci jest ciąg binarnych obrazów $Y[n]$ o rozmiarach równych obrazowi wejściowemu S . Na początku symulacji F i Θ są równe 0.

2.1. Sygnatury generowane przez sieć ICM

Sieć ICM (*Intersecting Cortical Model*) jest uproszczoną wersją sieci PCNN, zdefiniowaną przez Kinsera [7]. Uproszczenie polega na całkowitym pominięciu w modelu neuronu pętli sprzężenia zwrotnego L (*linking*) – por. rysunek 1. W tej sytuacji całkowite pobudzenie neuronu U jest równe F . Równania opisujące sieć ICM (zgodnie z rys. 1) przyjmują postać:

$$F_{ij}[n+1] = fF_{ij}[n] + S_{ij} + \sum_{kl} W_{ijkl} Y_{kl}[n] \quad (6)$$

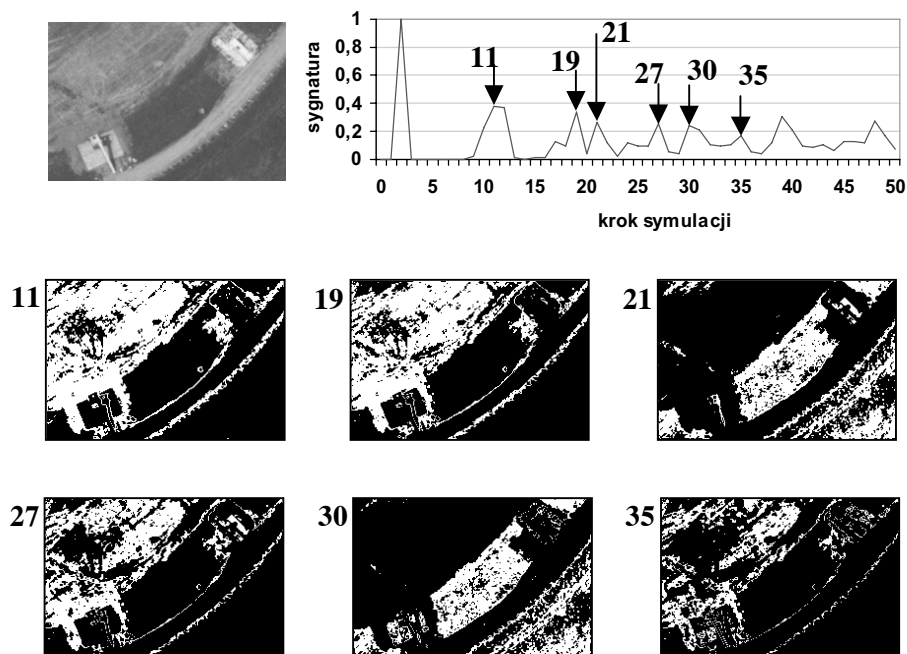
$$\Theta_{ij}[n+1] = g\Theta_{ij}[n] + hY_{ij}[n+1] \quad (7)$$

$$Y_{ij}[n+1] = \begin{cases} 1 & \text{gd}y \ F_{ij}[n+1] > \Theta_{ij}[n] \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad (8)$$

przy czym oznaczenia i zależności pomiędzy parametrami są jak poprzednio.

Sieci impulsujące generują sekwencje obrazów binarnych, na których w kolejnych krokach obliczeniowych uwypuklane są różne fragmenty obrazu wejściowego. W przypadku sieci ICM w zależności od wartości współczynników f , g i h oraz wag, definiujących wpływ neuronów sąsiednich (macierz W_{ijkl}) uzyskuje się różne efekty przetwarzania, np. wydzielanie krawędzi lub sekwencyjną detekcję obiektów. Dzieje się tak dzięki działaniu sprzężenia zwrotnego, które podtrzymuje wewnętrzne aktywności neuronów F .

Na rysunku 2 pokazano wyjścia sieci impulsującej ICM w wybranych krokach symulacji. Obszary białe odpowiadają neuronom generującym w danym kroku impuls. W pierwszym kroku (niepokazanym na rys. 2) prawie wszystkie neurony generują impuls, ponieważ dla obrazów rzeczywistych prawie wszystkie piksele są większe od zera – por. wzory (6) i (8), przy warunkach początkowych ($Y[0] = F[0] = \Theta[0] = 0$). Podczas kolejnego kroku „nie dzieje się nic”, ponieważ próg Θ wzrósł do wartości h przekraczającej wartości F . W kolejnych krokach próg stopniowo opada – por. wzór (7) i rysunek 1 – więc zaczynają impulsować neurony odpowiadające wysokim wartościom pikseli obrazu wejściowego oraz te, które otrzymują wysoki sumaryczny sygnał sprzężenia zwrotnego – por. wzór (6). W zależności od zadanych parametrów sieci rozkłady pobudzeń F różnicują się, powodując impulsacje różnych obszarów sieci. Obrazy impulsacji dla kilku kroków zostały zilustrowane na rysunku 2, przy czym do prezentacji wybrano obrazy, na których można zaobserwować różne wykrywane obszary.



Rys. 2. Obraz oryginalny, sygnatura 50-elementowa i obrazy rozkładu impulsacji z wybranych kroków symulacji sieci ICM
Objaśnienia w tekście

Koncepcja sygnatur obrazów (obiektów) została wprowadzona przez Johnsona w 1994 roku [9]. Podstawą obliczania sygnatur jest funkcja czasowa $G[n]$, otrzymywana w wyniku sumowania wyjść Y neuronów („białych pikseli”) w każdym kroku obliczeniowym

$$G[n] = \sum_{i,j} Y_{ij}[n] \quad (9)$$

Johnson stwierdził, że w przypadku analizy prostych obiektów znajdujących się na jednolitym tle po pewnej liczbie iteracji funkcja $G[n]$ staje się okresowa. Powtarzający się fragment został nazwany sygnaturą. Dla prostych obrazów pokazano, że kształt sygnatury jest niezależny od rotacji, przesunięcia, zmiany skali, a nawet zmiany kąta widzenia obiektu [4, 6, 9].

Zwykle rzeczywisty obraz nie zawiera prostych obiektów na czarnym tle, dlatego jako sygnaturę przyjmuje się przebieg funkcji G w pierwszych kilkudziesięciu krokach symulacji (przeważnie uwzględnia się 25 lub 50 kroków [1, 4]). W innych pracach [5] wskazuje się na fakt generowania przez sieć „cech”, obserwowanych jako lokalne maksima funkcji G i oblicza się zależność efektywności klasyfikacji od liczby uwzględnionych cech.

Na rysunku 2 przedstawiono przykład generacji 50-elementowej sygnatury z wycinka zdjęcia lotniczego. Na wykresie sygnatury zaznaczono kilka lokalnych maksimów, a niżej odpowiadające im obrazy impulsacji sieci. Widoczny w lewym górnym rogu obraz, dla którego obliczono sygnaturę jest obszarem korzystnym z punktu widzenia późniejszej analizy fotogrametrycznej, ponieważ znajdują się na nim kontrastowe obiekty, ułatwiające wybór punktów do późniejszego dopasowywania.

3. Klasyfikacja obrazów na podstawie sygnatur

Do utworzenia zbiorów uczącego i testowego dla sieci neuronowej wykorzystano dwa zdjęcia lotnicze okolic Krakowa, na których były widoczne zarówno obszary o zabudowie miejskiej oraz wiejskiej, jak i pola uprawne i lasy. Z pierwszego zdjęcia wyodrębniono 452 obszary o rozmiarach 241×161 pikseli i podzielono je na dwie klasy:

- 1) obszarów korzystnych pod względem późniejszego wykorzystania w analizie fotogrametrycznej (231 sztuk),
- 2) niekorzystnych (221 obrazów).

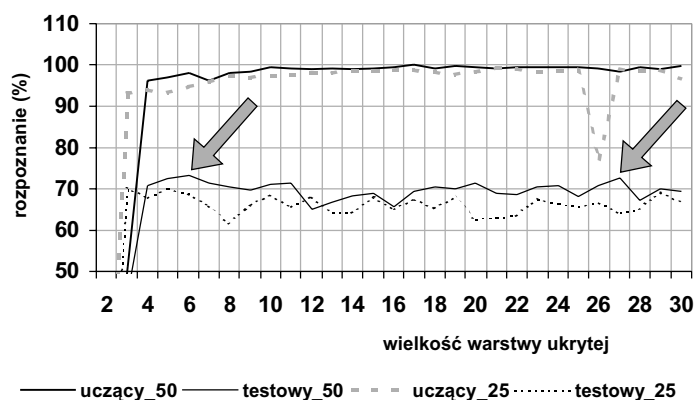
Podobnie wykorzystano drugie zdjęcie do utworzenia zbioru testowego: wybrano także 452 obrazy, z których 202 opisano jako obszary korzystne, a 250 jako niekorzystne z fotogrametrycznego punktu widzenia. W ten sposób (w przybliżeniu) zapewniono równoliczność obu klas obrazów.

Dla wszystkich obrazów wygenerowano – za pomocą sieci ICM – 50-elementowe sygnatury. Sygnatury znormalizowano do przedziału $[0, 1]$. Przykład obrazu źródłowego i jego sygnatury przedstawiono na rysunku 2.

3.1. Wybór wielkości warstwy ukrytej

Wielkość i liczba warstw ukrytych ma podstawowe znaczenie dla efektywności sieci typu *backpropagation* [11]. Badania rozpoczęto od sieci z jedną warstwą ukrytą. Do wstępnych eksperymentów wykorzystano Neural Network Toolbox [2], stanowiący część środowiska programistycznego Matlab. Podczas kolejnych eksperymentów zmieniano wielkość warstwy ukrytej od 3 do 30 elementów. Sieci były uczone metodą batchową przy użyciu algorytmu Levenberga–Marquardta.

Wyniki uczenia i testowania przedstawiono na rysunku 3.



Rys. 3. Wyniki uczenia i rozpoznawania zbiorów sygnatur 25- i 50-elementowych. Szare strzałki wskazują lokalne maksima rozpoznawania zbioru testowego dla sygnatur 50-elementowych

Tabela 1

Wyniki uczenia i rozpoznawania dla dwóch wielkości warstwy ukrytej

	Sieć 50-6-2*		Sieć 50-27-2*	
	zbiór uczący	zbiór testowy	zbiór uczący	zbiór testowy
Rozpoznanie	98,0%	73,23%	98,45%	72,79%
Macierz pomyłek	226 5 4 217	145 57 64 186	227 4 3 218	143 59 64 186

* Notacja: liczebność warstwy wejściowej-ukrytej-wyjściowej.

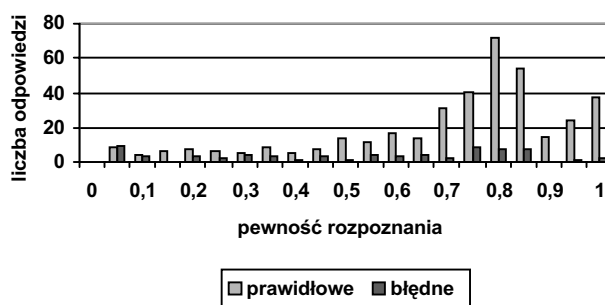
Z wykresu zamieszczonego na rysunku 3 wynika, że rezultaty uczenia są podobne dla sygnatur 25- i 50-elementowych, natomiast wyniki rozpoznawania są wyraźnie lepsze dla reprezentacji 50-elementowej. W tabeli 1 zestawiono w formie liczbowej wyniki dla dwóch wielkości warstwy ukrytej (6 i 27 neuronów), dla których rozpoznanie osiągnęło lokalne maksimum. Zarówno globalne rozpoznania, jak i macierze pomyłek są dla obu sieci bardzo zbliżone. Ze względu na potencjalnie większe możliwości generalizacji do dalszych eksperymentów wybrano sieć o sześciu neuronach w warstwie ukrytej.

3.2. Uczenie sieci 50-6-2 i korekta otrzymanych wyników

Automatyczna selekcja korzystnych (z punktu widzenia późniejszego dopasowywania) obszarów polega na sekwencyjnym pobieraniu ze zdjęcia lotniczego **kilkuset** obszarów o stałej wielkości, poddanie ich ocenie i wyborze **kilku** (kilkunastu) najlepszych. Z proporcji „kilkaset wejściowych/kilka wyjściowych” obrazów wynika, że można w tym przypadku zastosować technikę odrzuceń części odpowiedzi sieci, opierając się na analizie pewności rozpoznania. W przypadku dwóch rozpoznawanych klas pewnością rozpoznania będzie różnica pomiędzy wartościami wyjść sieci. Określony na podstawie analizy pewności rozpoznania próg odrzuceń stosuje się do zbioru testowego.

Opisany w podrozdziale 3.1 algorytm uczenia sieci doprowadza do bardzo dokładnego odwzorowania zbioru uczącego, co jest widoczne po obliczeniu pewności rozpoznania: w większości przypadków pewność ta jest równa 1, co oznacza, że jeden z neuronów warstwy wyjściowej generuje wartość 1, a drugi 0. Sytuacja ta uniemożliwia dokonanie korekty przez wprowadzenie progu pewności rozpoznania i odrzucenia części rozpoznań. Konieczne staje się przeprowadzenie powtórnego, stopniowego uczenia sieci i przerwania tego procesu na wcześniejszym etapie. Takie uczenie przeprowadzono metodą sekwencyjną przy użyciu oprogramowania NeuralWorks Professional II+ firmy NeuralWare. Program ten umożliwia obserwowanie na bieżąco nie tylko rozpoznań zbioru uczącego, ale także w każdym kroku aktualizuje i wyświetla macierz pomyłek. Uczenie zatrzymano, gdy rozpoznanie zbioru uczącego osiągnęło poziom około 85%, czyli było nieco wyższe od uzyskanego poprzednio rozpoznania zbioru testowego (73%).

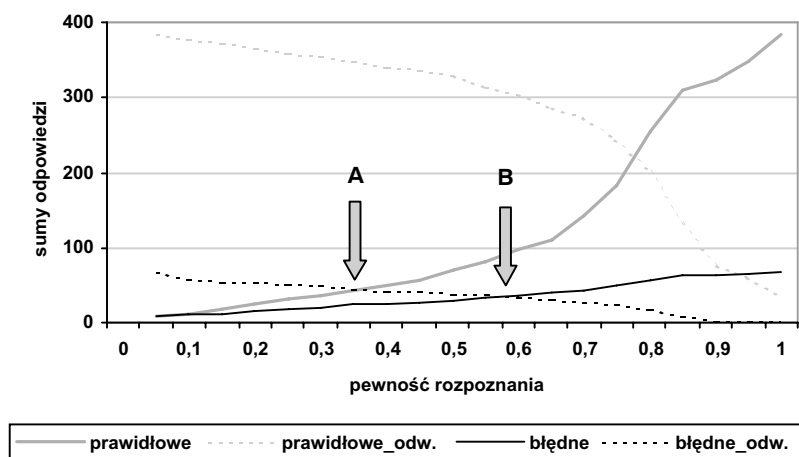
Dla tak nauczonej sieci obliczono histogram „pewności rozpoznania” (por. rys. 4), osobno dla odpowiedzi prawidłowych i błędnych. Z histogramu wynika, że znaczący wzrost liczby odpowiedzi prawidłowych ma miejsce w okolicy pewności rozpoznania równych 0,45 i 0,65. W tym drugim przypadku obserwuje się także wzrost liczby odpowiedzi błędnych: część błędnych odpowiedzi jest generowana przez sieć z dużą pewnością.



Rys. 4. Histogram pewności rozpoznań zbioru uczącego

Bardziej precyzyjne informacje można otrzymać analizując wykresy histogramów skumulowanych (por. rys. 5). Obliczono je na podstawie histogramów, przedstawionych na rysunku 4, sumując kolejne wartości w dwóch kierunkach: rosnących i malejących pewności rozpoznania. W ten sposób otrzymuje się kilka punktów charakterystycznych (przecięć

krzywych), które mogą ułatwić wybór progów odrzuceń. Na przykład ustalając próg akceptowalnej pewności w punkcie A powoduje się odrzucenie pewnej liczby odpowiedzi prawidłowych, a równocześnie zaakceptowanie takiej samej liczby odpowiedzi błędnych. Przyjęcie jako progów wartości B oznacza odrzucenie i równoczesne zaakceptowanie takiej samej liczby odpowiedzi błędnych. Dla progów równo 0,9 odrzucona jest większość danych, ale procent pozostałych rozpoznań zbliża się do 100.



Rys. 5. Histogramy pewności rozpoznania, kumulowane od wartości 0 do 1 i od 1 do 0 (odw.)

Progi zaproponowane na podstawie analizy histogramów pewności rozpoznania zbioru uczącego zostały zastosowane do otrzymanych wyników. Rezultaty dla zbioru uczącego i testowego zestawiono w tabeli 2; w zależności od przyjętego progów obserwuje się wzrost zarówno liczby odrzuceń, jak i rozpoznań (por. szare pola w tab. 2). Niestety w tym przypadku dla zbioru testowego nie sprawdza się trend, obserwowany w zbiorze uczącym: dla najwyższego progów pomimo wzrostu liczby odrzuceń rozpoznania zmalało. Oznacza to, że dane w zbiorze testowym znacząco różnią się od danych zbioru uczącego.

Tabela 2

Wyniki uczenia i rozpoznawania po zastosowaniu progów pewności rozpoznania

Próg pewności rozpoznania	Zbiór uczący		Zbiór testowy	
	odrzuconych	rozpoznanie	odrzuconych	rozpoznanie
0,0	–	85,17%	–	73,23%
0,4	16,37%	89,15%	20,57%	79,66%
0,6	29,86%	90,85%	40,48%	82,15%
0,9	85,84%	95,31%	84,51%	78,57%

4. Podsumowanie

W pracy sprawdzono możliwość użycia sygnatur obrazów do wstępnej selekcji fragmentów zdjęć lotniczych. Selekcja ma na celu określenie czy dany fragment zdjęcia lotniczego może być użyty w procesie dopasowywania kolejnych zdjęć.

Przetworzono ponad 900 fragmentów zdjęć lotniczych do postaci sygnatur 25- i 50-elementowych. Uzyskane w ten sposób dane podzielono na zbiór uczący i testowy. Na podstawie wstępnych eksperymentów wyznaczono wielkość warstwy ukrytej sieci typu *backpropagation* oraz zdecydowano, że dalsze eksperymenty będą prowadzone na zbiorach sygnatur 50-elementowych. Uzyskane rozpoznanie zbioru testowego na poziomie 73% świadczy o tym, że sygnatury mogą być użyte do wstępnej selekcji obrazów dla potrzeb fotogrametrii. Na zakończenie (por. podrozdz. 3.2) pokazano, w jaki sposób poprzez wprowadzenie progu pewności rozpoznania można – kosztem odrzucenia części danych – zwiększyć zarówno pewność rozpoznania jak i procentową skuteczność (do około 80%).

Dalsze badania powinny przebiegać w dwóch kierunkach. Po pierwsze należy zwiększyć zbiory danych, pozyskać obrazy z różnych serii zdjęć lotniczych i podczas podziału na obszary „korzystne” i „niekorzystne” uwzględnić opinie kilku ekspertów-fotogrametrów. Do weryfikacji wyników można użyć oprócz sieci typu *backpropagation* także sieci z rodziny SOM (*Self Organizing Maps*) Kohonena [11]. Drugim kierunkiem badań powinno być sprawdzenie, czy reprezentacja w postaci sygnatury może być użyta do innych zadań fotogrametrycznych, na przykład na etapie dopasowywania fragmentów zdjęć lotniczych z tej samej serii.

Literatura

- [1] Atmer J.: *Image Signatures from PCNN using Computers*. Dept. of Physics, Royal Institute of Technology (KTH), Stockholm 2003 (Diploma Work)
- [2] Demuth H., Beale M., Hagan M.: *Neural Network Toolbox 5 User's Guide*. The MathWorks, Inc., Natick, MA, 1992–2007
- [3] Eckhorn R., Reitboeck H.J., Arndt M., Dicke P.: *Feature Linking via Synchronisation among Distributed Assemblies: Simulations of Results from Cat Cortex*. *Neural Computation*, vol. 2, 1990, 293–307
- [4] Ekblad U., Kinser J.M., Atmer J., Zetterlund N.: *The intersecting cortical model in image processing*. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* Volume: 525, Issue: 1–2, June 1, 2004, 392–396
- [5] Forgáč R., Mokriš I.: *Pulse Coupled Neural Network Models for Dimension Reduction of Classification Space*. *Proc. Bratislava, WIKT 2006*
- [6] Johnson J.L., Padgett M.L.: *PCNN Models and Applications*. *IEEE Trans. on Neural Networks*, vol. 10, No 3, 1999, 480–498
- [7] Kinser J.M.: *A Simplified Pulse-Coupled Neural Network*. *Proc. SPIE*, vol. 2760, No. 3, 1996
- [8] Kurczyński Z.: *Lotnicze i satelitarne obrazowanie Ziemi*. Warszawa, Oficyna Wydawnicza Politechniki Warszawskiej 2006

-
- [9] Lindblad T., Kinser J.M.: *Image Processing Using Pulse-Coupled Neural Networks*. Heidelberg, Berlin, Springer-Verlag 2005
- [10] Mikrut Z., Tadeusiewicz R.: *Sieci neuronowe w przetwarzaniu i rozpoznawaniu obrazów*. [w:] *Biocybernetyka i Inżynieria Biomedyczna 2000*, t. 6 Sieci Neuronowe, AOW, Warszawa, EXIT 2000, 459–493
- [11] Tadeusiewicz R.: *Sieci neuronowe*. Warszawa, Akademicka Oficyna Wydawnicza RM 1993