



# ROZPOZNAWANIE ZNAKÓW POLSKIEGO ALFABETU PALCOWEGO

**Joanna Marnik**

*Politechnika Rzeszowska, Katedra Informatyki i Automatyki*

## Wprowadzenie

Interakcja człowiek–komputer staje się elementem naszej codzienności nabierającym coraz większego znaczenia. Można oczekiwać, że komunikacja ludzi z przyszłymi komputerami będzie się odbywać w naturalny sposób.

W ostatnich latach ważnym wyzwaniem dla twórców systemów komputerowych stała się budowa systemów rozpoznawania gestów. Ludzie posługują się gestami na co dzień w celu przekazywania między sobą informacji.

Gesty wykonywane przez człowieka dotyczą jego wyrazu twarzy, pozycji i, ewentualnie, ruchu całego ciała lub jego części, w szczególności rąk.

W oparciu o gesty wykonywane rękami powstały języki migowe, którymi posługują się osoby głuchonieme. W językach tych danemu wyrażeniu, słowu czy znakowi (literze, liczbie) odpowiada jednoznacznie określony gest. Stworzenie systemu wizyjnego tłumaczącego gesty języka migowego na język mówiony lub pisany umożliwiłoby porozumiewanie się osób niesłyszących z ludźmi nie związanymi ze środowiskiem osób z tego typu niesprawnością, w szczególności z urzędnikami, pracownikami banków i innych instytucji.

## Polski Język Migany (PJM) i Polski Alfabet Palcowy (PAP)

Osoby niesłyszące do porozumiewania się ze słyszącymi wykorzystują pewien charakterystyczny dla swojego kraju zestaw umownych znaków przekazywanych za pomocą rąk. W Polsce znaki te stanowią część kinestetyczną systemu językowo-migowego zwanego Polskim Językiem Migany (PJM). Polski Język Migany [4] oparty jest na gramatyce języka polskiego. Stosowany jest w sytuacjach „formalnych”, np. przy tłumaczeniu programów telewizyjnych, referatów na sympozjach, wypowiedzi na naradach i konferencjach, a także podczas rozmów osób niesłyszących ze słyszącymi, którzy posługują się tylko Polskim Językiem Migany.

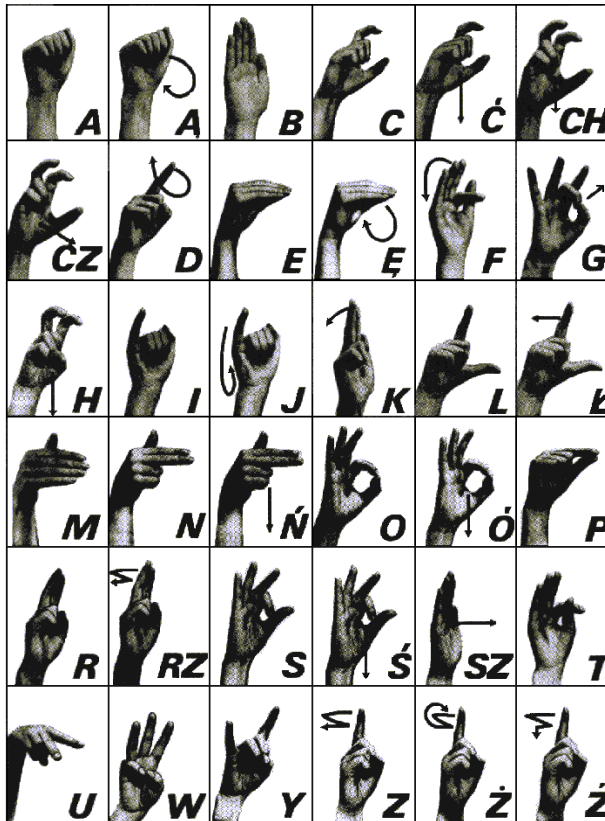
Znaki migowe PJM dzielą się na znaki ideograficzne i znaki daktylograficzne. Znaki ideograficzne są to znaki pojęciowe, oznaczające określone słowo lub zwrot. Podczas

przekazywania tych znaków ręka ustawiona jest w określonym miejscu względem ciała. Znaki określające litery alfabetu, liczby, działania arytmetyczne, skróty miar metrycznych i inne skrótowce to znaki daktylograficzne.

Znaki migowe przyporządkowane literom alfabetu polskiego oraz głoskom Ch, Cz, Rz i Sz składają się na tzw. Polski Alfabet Palcowy (PAP). Znaki te wykorzystywane są w Polskim Języku Miganym do:

- ♦ przekazywania imion, nazw, skrótowców oraz literowania wyrazów, dla których nie ma znaku ideograficznego,
- ♦ przekazywania formantów w wyrazach pochodnych i końcówek fleksyjnych.

W pierwszym przypadku ręka prezentująca znaki powinna znajdować się w odległości około 20 cm od twarzy, z prawej strony ust, natomiast w drugim na wysokości klatki piersiowej. Kompletny zestaw znaków Polskiego Alfabetu Palcowego przedstawiono na rys. 1.



Rys. 1. Zestaw znaków Polskiego Alfabetu Palcowego

Znaki PAP scharakteryzowane są przez układ dłoni i trajektorię ruchu dłoni przekazującej dany znak (na rys. 1 trajektorię oznaczono za pomocą strzałki określającej kształt i zakres ruchu). Znaki nieposiadające trajektorii to znaki statyczne, pozostałe to znaki dynamiczne.



Niektóre znaki dynamiczne są ruchomą wersją swoich odpowiedników statycznych, np. Ń jest ruchomą wersją znaku N.

## Cel pracy

Celem pracy było opracowanie i praktyczna weryfikacja metody rozpoznawania znaków Polskiego Alfabetu Palcowego, wchodzącego w skład Polskiego Języka Miganego, na podstawie obrazów z kamery kolorowej. Przyjęto nie wykorzystywać rękawic ułatwiających wyodrębnienie poszczególnych części dłoni, zakładając jednak, że gesty są wykonywane starannie. Intencją pracy było całościowe przedstawienie problematyki z uwzględnieniem oryginalnych rozwiązań. Zastosowane metody wiążą się zasadniczo z przetwarzaniem i rozpoznawaniem obrazów cyfrowych, z akcentem na morfologię matematyczną i wykorzystanie sieci neuronowych.

## Problemy związane z rozpoznawaniem znaków PAP

Dla rozpoznania danego znaku PAP konieczne jest określenie kształtu dłoni oraz stwierdzenie, czy przekazywany znak jest znakiem statycznym czy też dynamicznym. W przypadku znaków dynamicznych trzeba zazwyczaj określić także kształt trajektorii zakreślonej przez dłoń przekazującą znak.

Zadanie rozpoznawania znaków PAP wymaga więc rozwiązania następujących problemów:

- ♦ wykrycia dłoni na obrazie kolorowym zawierającym dłoń prezentującą znak PAP,
- ♦ rozpoznania kształtu dłoni,
- ♦ rozpoznania kształtu trajektorii ruchu dłoni.

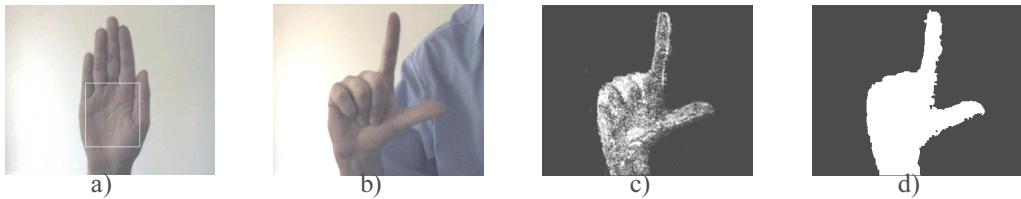
Poniżej omówiono krótko sposoby rozwiązania powyższych problemów. Szczegóły można znaleźć w pracy [6].

### *Wykrywanie dłoni*

Przebadano trzy metody wykrywania dłoni: metodę znaną z literatury, opartą na modelu barwy skóry ludzkiej (oznaczaną w pracy symbolem MBS) oraz dwie inne metody, opracowane w ramach pracy (MKR i MSN).

Metoda MBS polega na wydzieleniu z obrazu kolorowego pikseli odpowiadających dłoni w oparciu o model barwy skóry ludzkiej. Do stworzenia takiego modelu można wykorzystać jedną z wielu znanych przestrzeni barw [1, 5, 9]. W pracy przetestowano i porównano skuteczność metody MBS dla przestrzeni barw IHS, znormalizowanej przestrzeni RGB i przestrzeni YIQ, stosowanych najczęściej w literaturze. Testy wykazały, że najlepsze efekty daje użycie przestrzeni IHS, dla której otrzymano najmniejsze błędy w obszarze dłoni, bez względu na rodzaj oświetlenia.

W metodzie MBS model rozkładu koloru skóry ludzkiej generowany jest na podstawie wzorcowego obrazu, na którym ręcznie zaznaczano prostokątny obszar zawierający skórę (rys. 2 a). Utworzony model służy do przekształcenia obrazu, z którego należy wydzielić dłonie (rys. 2 b), do obrazu prawdopodobieństwa, na którym jasność piksela określa stopień jego przynależności do skóry (rys. 2 c). Obraz prawdopodobieństwa poddawany jest filtracji dolnoprzepustowej za pomocą maski Gaussa o wymiarach  $3 \times 3$  i progowaniu z wykorzystaniem metody opartej na aproksymacji histogramu za pomocą dwóch krzywych Gaussa [10]. W efekcie otrzymywany jest obraz binarny (rys. 2 d), na którym biały obszar określa miejsce występowania pikseli należących do dłoni.



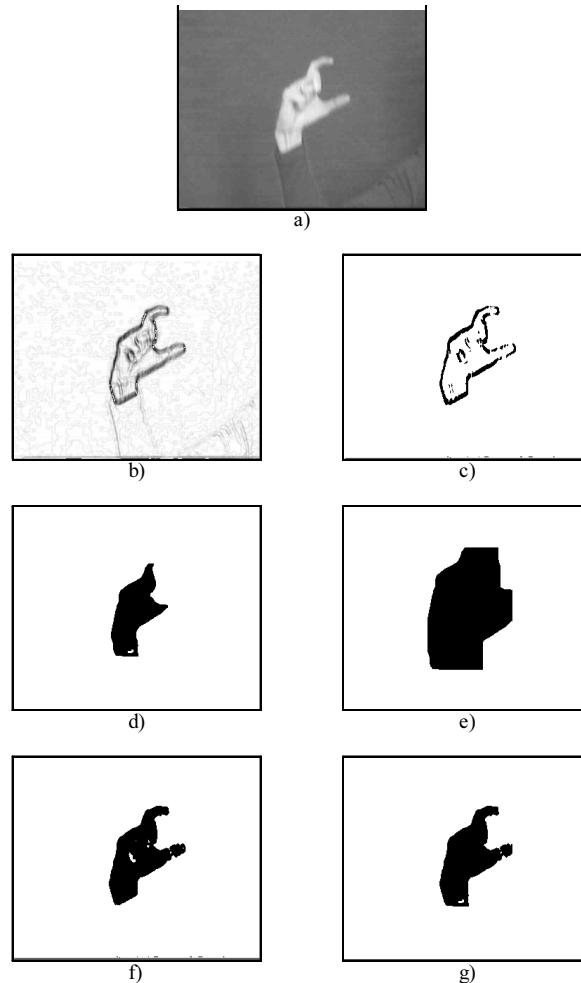
Rys. 2. Proces rozpoznawania skóry: a) fragment obrazu wykorzystywany do obliczenia modelu; b) analizowany obraz wejściowy; c) obraz prawdopodobieństwa; d) obraz c) po binaryzacji

Obrazy binarne otrzymywane metodą MBS często zawierały dłonie, w której występowały spore ubytki, ujawniające się zazwyczaj jako wybrakowane końcówki palców. Dlatego opracowano metodę MKR. Metoda ta ma na celu poprawę klasyfikacji uzyskiwanej metodą MBS, dzięki wykorzystaniu informacji o krawędziach. Obraz wynikowy metody MKR (*ObrazMKR*) powstaje w oparciu o następujące obrazy:

1. obraz obszarów odpowiadających dłoni (*ObszarDłoni*),
2. obraz powiększonej dłoni z obrazu wynikowego metody MBS (*DłońPowiększona*),
3. obraz wynikowy metody MBS (*ObrazMBS*),

poprzez wyznaczenie iloczynu logicznego obrazów *DłońPowiększona* i *ObszarDłoni*, a następnie sumy logicznej powstałego obrazu i obrazu *ObrazMBS*. Obraz *ObszarDłoni* powstaje poprzez wypełnienie obszaru wyznaczonego przez krawędzie dłoni obecnej w analizowanym obrazie szarym, przy zastosowaniu algorytmu wypełniania obszarów na podstawie fragmentów brzegu [6, 10]. Do wyznaczenia obrazu krawędzi stosowano operator Sobela. Obraz *DłońPowiększona* tworzony był poprzez 15-krotne wykonanie morfologicznej operacji dylatacji [6, 8, 12] na obrazie *ObrazMBS*, z użyciem elementu strukturalnego o wymiarach  $3 \times 3$ . Przykładowy obraz otrzymany metodą MKR i obrazy pośrednie, z których powstał, pokazano na rys. 3.

Zamierzone skutki działania metody można osiągnąć jedynie wtedy, gdy spełnione są warunki jednorodności tła i równomierności oświetlenia oraz wyraźnej różnicy poziomów szarości dłoni i tła.



Rys. 3. Ilustracja działania metody MKR: a) obraz dłoni; b) obraz krawędzi; c) obraz krawędzi po progowaniu; d) obraz binarny otrzymany metodą MBS; e) obraz d) po dylatacji; f) obraz dłoni powstały przez wypełnienie obszaru dłoni w oparciu o obraz krawędzi; g) obraz otrzymany metodą MKR

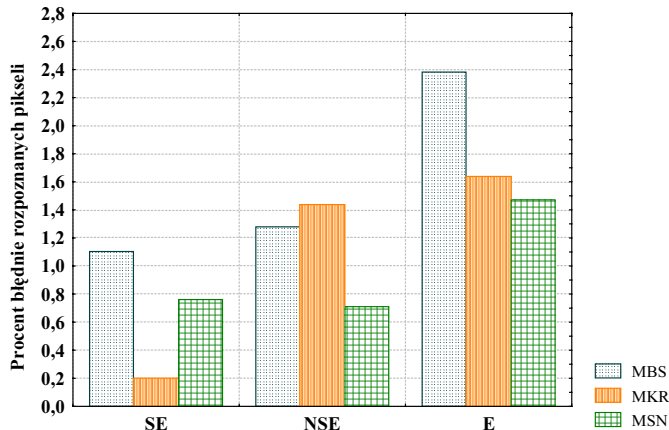
Uzyskiwanie obrazów binarnych dłoni metodą MBS o jakości nie w pełni zadowalającej pod względem możliwości rozpoznawania kształtu dłoni oraz ograniczenia, jakie ma metoda MKR doprowadziły do powstania metody MSN. W metodzie tej do wydzielania dłoni wykorzystano trójwarstwową, jednokierunkową sieć neuronową o wejściach odpowiadających składowym R, G i B barwy analizowanego piksela obrazu i wyjściu wskazującym, czy piksel ten należy do dłoni czy do tła. Warstwa ukryta tej sieci składała się z 16 neuronów z logistyczną funkcją aktywacji. Struktura sieci zaprojektowana została z użyciem narzędzia *Intelligent Problem Solver* (IPS) pakietu *STATISTICA Neural Networks (SNN)* [11]. Sieć o strukturze ustalonej w wyniku działania procedury IPS dopasowywana była do warunków, w jakich pozyskiwano obrazy dłoni z wykorzystaniem

algorytmu Levenberga-Marquardta. Przykładowe obrazy binarne dłoni uzyskane metodą MSN przedstawiono na rys. 4.



Rys. 4. Przykładowe obrazy otrzymane metodą MSN

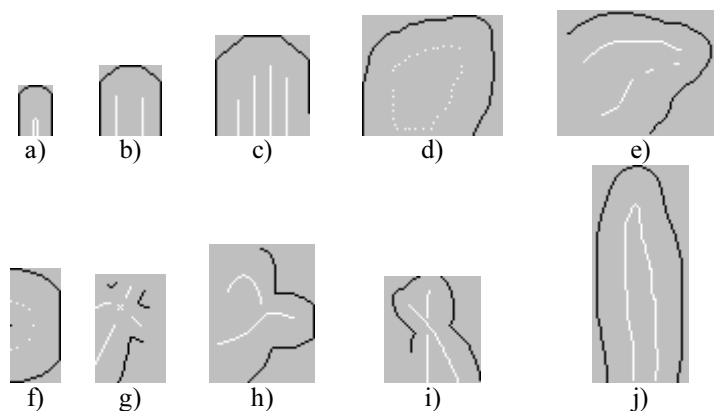
Testy porównawcze metod MBS, MKR i MSN wykazały, że najlepsze rezultaty daje metoda MSN, dla której uzyskiwano najmniejsze wartości błędów w obszarze tła, przy jednoczesnym dobrym odwzorowywaniu dłoni. Otrzymywanie niewielkich błędów w tle wynika z uwzględnienia w procesie uczenia sieci elementów tła obrazów testowych. Zbliżone do metody MSN efekty otrzymano dla metody MKR, jednakże metoda ta, jak wspomniano wcześniej, nie może być stosowana w każdych warunkach. Wyniki testów otrzymane dla 192 obrazów dłoni o rozdzielczości  $352 \times 288$ , pozyskanych w różnych warunkach oświetlenia, przedstawiono na rys. 5.



Rys. 5. Porównanie metod wydzielenia dłoni (SE – (*Skin Error*) procent pikseli obrazu należących do skóry zakwalifikowanych do tła; NSE – (*Non Skin Error*) procent pikseli obrazu należących do tła sklasyfikowanych jako piksele skóry; E = SE + NSE)

## Rozpoznawanie kształtu dłoni

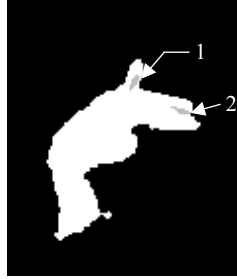
W celu rozwiązania zadania rozpoznawania kształtu dłoni opracowano metodę opartą na morfologicznym przekształceniu trafi–nie trafi, służącym do wykrywania cech geometrycznych obiektów binarnych [6, 7]. Zasada działania tej transformacji jest następująca [6, 8, 12]. Element strukturalny stanowiący podstawę operacji składa się z dwóch rozłącznych elementów B1 i B2 i ma wyróżniony punkt odniesienia  $\Omega$ . Punkt ten jest przesuwany przez wszystkie piksele obrazu. W wyniku transformacji wartość 1 otrzymują piksele, dla których po umieszczeniu w nich punktu  $\Omega$ , element B1 w całości znajduje się w obiekcie, element zaś B2 w tle. Pozostałym pikselom zostaje przypisana wartość 0.



Rys. 6. Elementy strukturalne wykorzystywane przy analizie układu dłoni służące do wykrywania: a) pojedynczego palca; b) dwóch złączonych palców; c) czterech złączonych palców; d) dłoni z zaciśniętymi palcami; e) kształtu charakterystycznego dla znaków E i P; f) złączonych końcówek kciuka i palca wskazującego (układ palców dla znaków O i G); g) kciuka skrzyżowanego z palcem wskazującym (układ palców dla znaków F i T); h) palca wskazującego złączony z kciukiem w jego środkowej części (układ palców dla znaku S); i) skrzyżowanych palców środkowego i wskazującego (układ palców dla znaku R); j) układu dłoni dla znaku Sz

Dla analizy kształtu dłoni podstawowe znaczenie ma utworzenie bazy elementów strukturalnych. Elementy te przedstawiono na rys. 6. Każdy z nich jest przeznaczony do detekcji innego fragmentu dłoni. Kolorem białym zaznaczone są części B1 elementów strukturalnych, które powinny zawierać się w obiekcie (trafi), kolorem czarnym części B2, które powinny znajdować się w tle (nie trafi), natomiast kolorem szarym obszary, w których wartość pikseli analizowanego obrazu nie ma wpływu na wynik transformacji. Transformacja trafi–nie trafi z zadaniem elementem strukturalnym wykrywa elementy obrazu o postaci zakodowanej w tym elemencie. Stąd np. wykrycie dwóch palców ułożonych poziomo przy użyciu elementu strukturalnego z rys. 6 b nie będzie możliwe, jeśli nie zostanie dokonany jego obrót o kąt  $90^\circ$ . Dla rozróżnienia symboli Polskiego Alfabetu Palcowego do bazy elementów strukturalnych dołączono więc wszystkie wymienione elementy strukturalne obrócone o odpowiednie kąty. Zakresy obrotu i krok, z jakim zmieniano kąt obrotu, zostały podane w pracy [6].

Kontury wykrywanych fragmentów dłoni mogą się zmieniać w zakresie określonym przez obszar znajdujący się pomiędzy częściami B1 i B2 elementów strukturalnych. Zdefiniowanie takiego obszaru „neutralnego” jest konieczne ze względu na różnice w proporcjach pomiędzy poszczególnymi fragmentami dłoni u różnych osób, wahania odległości od kamery oraz błędy segmentacji. Przykład wykrycia pojedynczego palca skierowanego do góry i dwóch złączonych palców skierowanych w bok podczas prezentacji znaku N pokazano na rys. 7.

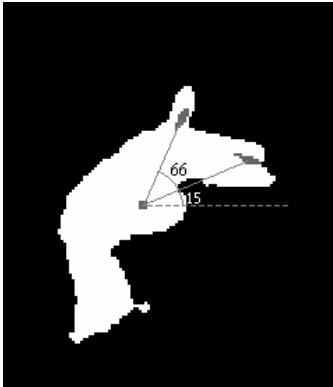


Rys. 7. Obiekty wynikowe transformacji trafi–nie trafi powstałe przy użyciu: 1 – elementu strukturalnego z rys. 6 a; 2 – elementu strukturalnego z rys. 6 b obróconego o kąt  $90^\circ$

Poza kształtem zakodowanym w elemencie strukturalnym ważny jest także jego rozmiar. Dlatego konieczne jest dopasowanie wielkości obiektu znajdującego się na obrazie i kształtów opisanych za pomocą elementów strukturalnych. Przyjęto, że skalowaniu poddawany będzie obraz dłoni. Wartość skali ustalana jest na podstawie informacji o długości średnicy maksymalnego koła zawartego całkowicie w obiekcie reprezentującym dłoń z wyprostowanymi wszystkimi palcami i wewnętrzną częścią zwróconą w kierunku kamery. Obraz tak ułożonej dłoni prezentowany jest przed przystąpieniem do prezentacji znaków PAP. Szczegóły rozwiązania przedstawiono w pracy [6].

Kolejnym krokiem w procesie rozpoznawania kształtu dłoni jest formowanie wektora cech. W wektorze cech są zapisywane informacje o fragmentach dłoni wykrytych metodą trafi–nie trafi. Analizie podlegają wszystkie obiekty powstałe w wyniku transformacji. Dla każdego takiego obiektu są wyznaczane następujące parametry: (i) powierzchnia, (ii) kąt nachylenia półprostej mającej początek w środku ciężkości dłoni i przechodzącej przez środek ciężkości analizowanego obiektu do poziomu, (iii) odległość środka ciężkości analizowanego obiektu od środka ciężkości dłoni. Powyższe parametry wraz z kodem liczbowym elementu strukturalnego, z użyciem którego powstał obiekt im odpowiadający, stanowią czwórkę liczb opisujących pewien fragment dłoni. Wektor cech składa się z pięciu takich czwórek zapisanych w kolejności malejących kodów. Kod liczbowy przyporządkowany określonemu elementowi strukturalnemu odzwierciedla wagę kształtu reprezentowanego przez ten element w zestawie znaków PAP. Przykładowy wektor cech oraz sposób jego wyznaczania dla znaku N z rys. 7 przedstawiono na rys. 8. W wektorze tym wypełnione są tylko dwie czwórki określające kształt. Pierwsza z nich opisuje obiekt powstały w wyniku wykrycia dwóch złączonych palców skierowanych w bok. Kod elementu strukturalnego charakteryzującego taki kształt wynosi 5. Obiekt reprezentowany jest przez 19 pikseli. Odcinek łączący środki ciężkości tego obiektu i dłoni jest nachylony

do poziomu pod kątem  $15^\circ$ , a jego długość wynosi 45 jednostek. Obiekt opisujący pojedynczy palec skierowany w górę został wykryty przez element strukturalny o kodzie 1. Kolejne trzy liczby związane z tym kodem należy interpretować jak poprzednio. Trzy pozostałe czwórki są wypełnione wartościami  $-1$ , co oznacza, że nie opisują żadnego obiektu wynikowego transformacji trafi–nie trafi.



Wektor cech:

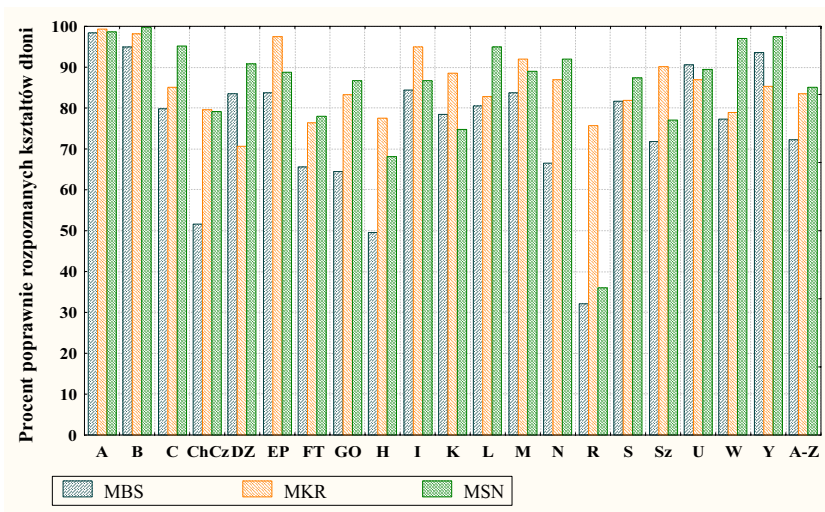
```
(5 19 15 45 1 21 66 34 -1 -1 -1 -1  
-1 -1 -1 -1 -1 -1 -1 -1)
```

Rys. 8. Wektor cech dla symbolu znaku N z rys. 7

Jak widać z rys. 1, niektóre litery PAP mają identyczny układ dłoni, różniąc się jedynie występowaniem i kształtem ruchu. Tak więc rozpoznawano 20 następujących klas kształtu dłoni (obok przyjętej nazwy klasy w razie potrzeby podano w nawiasie odpowiadające jej litery): A(A, A), B, C(C, C), ChCz(CH, CZ), DZ(D, Z, Ż, Ż), EP(E, E, P), FT(F, T), GO(G, O, Ó), H, I(I, J), K, L, M, N(N, N), R(R, RZ), S(S, S), SZ, U, W, Y.

Struktura sieci neuronowej służącej do rozpoznawania kształtu dłoni została zaprojektowana z użyciem narzędzia *Intelligent Problem Solver* (IPS) pakietu *STATISTICA Neural Networks* (SNN) [11]. Testowano na wstępie sieci perceptronowe z jedną i dwiema warstwami ukrytymi. Przy wyborze parametrów działania procedury IPS zaznaczono opcję uruchamiającą narzędzie automatycznego wyboru zmiennych wejściowych. Wejście sieci stanowiły wektory cech utworzone w oparciu o wyniki transformacji trafi–nie trafi wygenerowane dla wybranych obrazów pochodzących z 8 filmów, na których prezentowany był kompletny zestaw znaków Polskiego Alfabetu Palcowego. Liczba elementów każdej klasy w zbiorze uczącym wynosiła około 300. Część wektorów cech zbioru uczącego została przeznaczona do weryfikacji poprawności uczenia sieci (w SNN, tzw. cross verification). Dzięki weryfikacji możliwe jest stwierdzenie, czy nie dochodzi do tzw. przecuczenia sieci, co wiąże się z utratą zdolności sieci do uogólniania zdobytej wiedzy. Rozmiar zbioru uczącego wynosił 5361 elementów, natomiast zbioru walidacyjnego 500 elementów. Dodatkowo wyodrębniono 500-elementowy zbiór testowy, dzięki któremu można upewnić się, czy otrzymana sieć działa poprawnie. Wyjście sieci stanowiło 20 neuronów, co odpowiada liczbie klas, które należy rozróżniać. Na podstawie szczegółowej analizy wrażliwości, którą umożliwia opcja *Sensitivity Analysis* pakietu SNN pominięto elementy wektora wejściowego, których wpływ na skuteczność rozpoznawania okazał się niewielki. Ostatecznie przyjęto do dalszych badań jedenastoelementowy wektor

cech. Dla uproszczonego wektora cech zaprojektowano jednokierunkową sieć perceptronalną o 11 neuronach wejściowych, 25 neuronach w warstwie ukrytej i 20 neuronach wyjściowych. Neurony warstw ukrytej i wyjściowej miały logistyczną funkcję aktywacji. Sieć była uczona metodą wstecznej propagacji błędów ze zmiennym współczynnikiem uczenia przez 50 epok, a następnie metodą gradientów sprzężonych przez 52 epoki. Projektując sieć, wykorzystano zbiory scharakteryzowane poprzednio. Skuteczność rozpoznawania przebadano następnie dla bazy danych liczącej łącznie 11661 obrazów z 8 filmów, o których była mowa wyżej. Wśród tych danych było też wspomniane 6361 obrazów z etapu konstruowania sieci. Wyniki uzyskane dla obrazów otrzymanych metodami MBS, MKR i MSN przedstawiono na rys. 9.



Rys. 9. Wyniki rozpoznawania układu dłoni

Ze względu na duże podobieństwo obrazów binarnych odpowiadających układom dłoni dla znaków A, E i P Polskiego Alfabetu Palcowego konieczne było zastosowanie dodatkowej metody, służącej do rozróżniania obiektów o podobnym kształcie w oparciu o obraz krawędzi dłoni. Metoda ta polega na utworzeniu histogramu orientacji [2] dla danego znaku. Histogram orientacji klasyfikowano z użyciem perceptronalnej sieci neuronowej o 36 neuronach wejściowych, 12 neuronach z logistyczną funkcją aktywacji w warstwie ukrytej i 3 neuronach z logistyczną funkcją aktywacji w warstwie wyjściowej. Neurony warstwy wyjściowej odpowiadały trzem klasom rozróżnianych z użyciem tej sieci kształtów dłoni (A, E, P). Sieć została zaprojektowana z użyciem narzędzia *Intelligent Problem Solver* pakietu *STATISTICA Neural Networks*. Histogramy orientacji stanowiące zbiór uczący otrzymano na bazie 8311 obrazów znaków A, E i P, przygotowanych specjalnie dla celów uczenia sieci. Do uczenia zastosowano metodę wstecznej propagacji błędów (50 epok), a następnie metodę gradientów sprzężonych (52 epoki). Wyniki rozpoznawania znaków A, E i P pochodzących z filmów, dla których otrzymano wyniki dla metody rozpoznawania kształtu dłoni wykorzystującej transformację trafi–nie trafi, przedstawione na rys. 9, zestawiono w tab. 1.

Tabela 1. Wyniki rozpoznawania znaków A, E i P z wykorzystaniem histogramów orientacji

|                        | A    | E    | P    |
|------------------------|------|------|------|
| Liczba obrazów         | 1014 | 1201 | 379  |
| Błędnie sklasyfikowane | 20   | 47   | 5    |
| Błąd w procentach      | 1,97 | 3,91 | 1,32 |

Zaproponowana w pracy metoda rozpoznawania kształtu dłoni, wykorzystująca transformację trafi–nie trafi, histogramy orientacji i sieci neuronowe stanowi oryginalne rozwiązanie, nie stosowane dotychczas w systemach rozpoznawania gestów. Wyniki jej zastosowania do 11 661 obrazów z 8 wspomnianych wcześniej filmów przedstawiono w tab. 2.

Tab. 2. Wyniki rozpoznawania kształtu dłoni znaków PAP

| Osoby wykonujące gesty                       | Metoda wydzielania dłoni |      |      |
|--|--------------------------|------|------|
|  | MBS                      | MKR  | MSN  |
| Autorka pracy                                | 75,1                     | 87,5 | 87,6 |
| Osoby związane ze środowiskiem niesłyszących | 80,1                     | 68,7 | 77,7 |

Proces opracowywania metody oparty był na gestach wykonywanych przez autorkę pracy. Wyniki zamieszczone w tabeli dotyczą gestów niewykorzystywanych w tym procesie. Niska skuteczność otrzymana dla metody MKR, w przypadku gestów wykonywanych przez osoby związane ze środowiskiem niesłyszących, jest wynikiem niespełnienia założenia tej metody, dotyczącego równomierności oświetlenia dłoni, przy nagrywaniu filmów objętych analizą. Największe błędy rozpoznania otrzymywano dla układów dłoni charakterystycznych dla znaków H, C, Ch. Są one wynikiem przestrzennego charakteru tych znaków. Pewność rozpoznania tych znaków może zostać zwiększona przy zastosowaniu metod opartych na obrazach otrzymanych z dwóch kamer. Problemy występowały także dla kształtu dłoni otrzymanego metodą MSN, odpowiadającego znakowi R, który był mylony ze znakami D i Z. Problemy te są wynikiem dużego podobieństwa tych znaków w Polskim Alfabcie Palcowym. Metoda opisu kształtu dłoni oparta na transformacji trafi–nie trafi posiada jednak szereg zalet. Oto one:

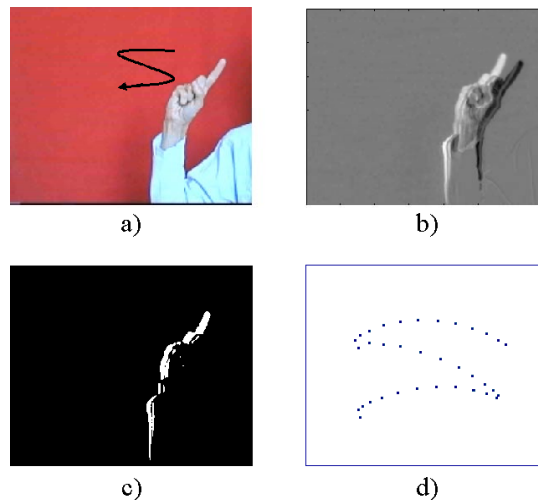
- ◆ niewrażliwość na niewielkie zakłócenia w tle i drobne niedoskonałości w odzwierciedleniu dłoni w postaci obrazu binarnego,
- ◆ brak konieczności odcinania fragmentów obrazu innych niż dłonie,
- ◆ możliwość zaadoptowania metody do gestów innych niż stanowiące przedmiot prezentowanej pracy.

Wady metody są następujące:

- ◆ konieczność ręcznego doboru elementów strukturalnych,
- ◆ występowanie trudności dla znaków posiadających przestrzenny charakter.

## Rozpoznawanie gestów dynamicznych

Trajektoria zakreślana przez dłoń podczas przekazywania dynamicznych symboli PAP może przybrać jeden następujących kształtów: linia pozioma skierowana w prawo (znak Sz), linia pozioma skierowana w lewo (znaki K i Ł), linia pionowa skierowana w dół (znaki Ć, Ch, H, Ń, Ó i Ś), wycinek okręgu zakreślany w kierunku zgodnym z ruchem wskazówek zegara (znaki A, E i D), trajektoria dla znaków F i J, łamana o kształcie litery Z (znaki Z i Rz), trajektoria znaku Ż, trajektoria znaku Ż, trajektorie dla znaków Cz i G. Do wyznaczania trajektorii wykorzystano metodę przepływu optycznego [3]. Z jej użyciem, na bazie kolejnych trzech obrazów z sekwencji reprezentującej ruch, wyznaczane są obszary, w których wystąpił ruch (tzw. pole energii ruchu). Obszary takie można przedstawić w postaci obrazu, na którym białe piksele odpowiadają miejscom, w którym prędkość ruchu przekroczyła ustalony próg (rys. 10 c).

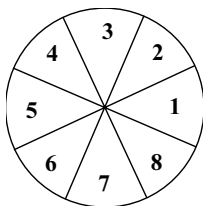


Rys. 10. Wyznaczanie trajektorii ruchu dłoni: a) znak Z; b) obraz pola energii ruchu; c) pole energii ruchu po binaryzacji; d) otrzymana trajektoria

Na podstawie ciągu obrazów pola energii ruchu generowana jest trajektoria zakreślona przez dłoń. Kolejne punkty trajektorii odpowiadają najwyższym położonym pikselom pola energii ruchu. Przykładową trajektorię otrzymaną dla znaku Z pokazano na rys. 10 d. Trajektoria opisywana jest za pomocą ciągu liczb, z których każda określa kierunek zmiany położenia dłoni w kolejnych chwilach czasu, kodowany z użyciem liczb, w sposób pokazany na rys. 11.

Zakodowana trajektoria z rys. 10 d ma postać:

[4 5 5 5 5 5 5 5 6 6 6 8 8 3 1 1 1 1 1 8 8 8 7 3 5 5 5 5 5 6 6 6 8]



Rys. 11. Sposób kodowania kierunków

Po zakodowaniu trajektoria poddawana jest filtracji mającej na celu usunięcie zakłóceń pojawiających się zazwyczaj wtedy, gdy dłoń zmienia gwałtownie kierunek ruchu lub porusza się zbyt wolno. Zakłócenia takie mają zazwyczaj postać punktów położonych daleko od trajektorii. W większości przypadków punkt taki reprezentowany jest przez dwa kolejne elementy odpowiadające przeciwnym kierunkom. Filtracja polega na usunięciu takich punktów. Trajektoria z rys. 10 d po filtracji przybiera postać:

[4 5 5 5 5 5 5 5 6 6 6 8 8 3 1 1 1 1 1 8 8 8 5 5 5 5 5 6 6 6 8]

Trajektorie klasyfikowano z użyciem sieci neuronowej. W tym celu poddawano je przeskalowaniu z użyciem interpolacji metodą najbliższego sąsiada (zgodnie z algorytmem użytym w funkcji *imresize* pakietu MATLAB 4.2c) do długości 20 elementów. Przykładowa trajektoria z rys. 10 d po przeskalowaniu przyjmuje postać:

[4 5 5 5 5 6 6 8 3 1 1 1 8 8 5 5 5 6 6 8]

Tak otrzymany ciąg stanowi wektor cech podawany na wejście perceptronowej sieci neuronowej zaprojektowanej z użyciem narzędzia *IPS* pakietu *SNN*. Sieć ta posiada 20 wejść, 16 neuronów o logistycznej funkcji aktywacji w warstwie ukrytej i 8 neuronów z logistyczną funkcją aktywacji w warstwie ukrytej. Liczba wyjść sieci odpowiada 8 pierwszym klasom kształtów trajektorii wymienionym wyżej. Ze względu na różnorodność kształtów trajektorii otrzymywanych dla znaków Cz i G, wynikającą z ich przestrzennego charakteru, trajektorii tych nie uwzględniano przy klasyfikacji. Jest to dopuszczalne, ponieważ stwierdzenie, że dłoń poruszała się podczas przekazywania znaku i zakreślona trajektoria ma kształt różny od linii skierowanej w dół, jednoznacznie wskazuje, że przekazywany znak to G lub Cz. Do uczenia sieci wykorzystano 2004 wektory cech odpowiadające trajektoriom otrzymanym dla dynamicznych znaków PAP, przygotowanych specjalnie dla celów uczenia sieci. Ze zbioru tego wydzielono 200-elementowy zbiór walidacyjny. Sieć uczona była najpierw metodą wstecznej propagacji błędów (50 epok), a następnie metodą gradientów sprzężonych (393 epoki).

Skuteczność metody przebadano dla trajektorii znaków dynamicznych pochodzących z 8 zestawów znaków PAP, o których była mowa wyżej, spośród których 6 zestawów prezentowanych było przez autorkę pracy (grupa 1), a 2 zestawy przez osoby związane ze środowiskiem niesłyszących (grupa 2). Poprawnie sklasyfikowanych zostało 97,2% trajektorii z grupy 1 i 92,2% trajektorii z grupy 2.



Zalety zaproponowanej metody rozpoznawania kształtów trajektorii są następujące:

- ◆ niewrażliwość na umiejscowienie trajektorii w obrazie,
- ◆ niewrażliwość na wielkość ruchu,
- ◆ możliwość adaptacji metody dla innych zadań, w których konieczna jest znajomość kształtu trajektorii,
- ◆ niezależność od częstotliwości pobierania kolejnych obrazów z kamery.

Informacja podana w ostatnim punkcie nie może być stosowana bezkrytycznie. Zakłada się bowiem, że czas trwania ruchu dla danego gestu obejmował będzie co najmniej 5 klatek. Metoda posiada następujące wady, które mogą mieć znaczenie w pewnych zastosowaniach:

- ◆ ze względu na przeskalowanie trajektorii do określonego rozmiaru dla potrzeb sieci neuronowej następuje utrata informacji o czasie trwania gestu,
- ◆ w opisie trajektorii nie jest uwzględniana szybkość ruchu.

Wymienione wyżej wady wymagają wprowadzenia drobnych modyfikacji do zaproponowanej metody w zastosowaniach, w których istotna jest informacja o czasie trwania ruchu i szybkości poruszania się dłoni.

## Osiągnięte rezultaty

Z użyciem zaproponowanych metod rozpoznawano znaki Polskiego Alfabetu Palcowego ze skutecznością ok. 94 %, jeśli gesty pokazywane były przez jedną osobę i ok. 80%, gdy wykonywały je różne osoby.

Opracowane w pracy metody zostały przetestowane pod względem szybkości obliczeń na multikomputerze złożonym z czterech komputerów z procesorami Intel Celeron 466 MHz nadzorowanych przez serwer z procesorem Intel Celeron 366 MHz. Dla obrazów kolorowych 352×288 uzyskano możliwość przetwarzania z częstotliwością 5 klatek/s.

## Podsumowanie

Zaproponowane rozwiązania mogą być z powodzeniem stosowane dla innego zestawu gestów niż założony w pracy, np. gestów komunikacji z komputerem lub robotem. Baza elementów strukturalnych, zbudowana dla celów rozpoznawania znaków Polskiego Alfabetu Palcowego, ma bowiem charakter bardziej ogólny i może być pomocna także przy rozpoznawaniu innych układów dłoni. Jednocześnie wydaje się, że kształty dłoni obecne w Polskim Alfabetcie Palcowym są na tyle skomplikowane, że trudno jest uformować palce tak, by uzyskać układ, któremu nie odpowiadałby żaden z elementów strukturalnych z zaprojektowanej bazy.



Zbudowanie systemu wizyjnego tłumaczącego gesty Polskiego Języka Miganego na język mówiony lub pisany umożliwiłoby porozumiewanie się osób niesłyszących z osobami niedotkniętymi tego typu niesprawnością, w szczególności z urzędnikami, pracownikami banków i innych instytucji. Powstałby równocześnie zbiór metod i narzędzi o ogólniejszym zastosowaniu w komunikacji człowieka z komputerem (robotem), np. w zadaniach sterowania urządzeniami. Przedstawiona praca stanowi krok w tym kierunku.

## Wykorzystanie programów z rodziny *STATISTICA*

W pracy wykorzystano następujące narzędzia z rodziny *STATISTICA*: program *STATISTICA Neural Networks* oraz program *STATISTICA*.

Pierwszy z wymienionych programów zastosowano do skonstruowania sieci neuronowych przeznaczonych do wydzielania dłoni z obrazu kolorowego, rozpoznawania kształtu dłoni oraz klasyfikacji kształtu trajektorii ruchu dłoni prezentującej dynamiczny znak PAP. Najczęściej stosowanymi funkcjami programu były:

- ◆ narzędzie *Intelligent Problem Solver*, z użyciem którego projektowano wstępną strukturę sieci neuronowych,
- ◆ funkcje związane z samodzielnym konstruowaniem sieci typu MLP,
- ◆ analiza wrażliwości sieci na zmienne wejściowe,
- ◆ funkcja *Classification Statistics* wykorzystana do wygenerowania wyników dla danych testowych.

Wyniki klasyfikacji kształtów dłoni (opcja *Classification Statistics* programu *STATISTICA Neural Networks*) uzyskane dla filmów zawierających gesty PAP wchodzące w skład przygotowanej w ramach pracy bazy danych, łączone były w wyniki zbiorcze z wykorzystaniem programów przygotowanych w języku *STATISTICA BASIC*, dostępnym z poziomu programu *STATISTICA*. Dane uzyskane w ten sposób posłużyły do wygenerowania w programie *STATISTICA* wykresów obrazujących wyniki testów. Danych tych użyto także do utworzenia w programie *MS Excel* tabel z wynikami rozpoznawania kształtów dłoni. Wykorzystano tu możliwość przenoszenia danych z programu *STATISTICA* do programu *MS Excel*.

## Literatura

1. Domański M., *Zaawansowane techniki kompresji obrazów i sekwencji wizyjnych*, Wyd. Politechniki Poznańskiej, Poznań, 2000.
2. Freeman W. T., Roth M., *Orientation histograms for hand gesture recognition* [w:] Proc. Of the IEEE Intl. Wkshp. On Automatic Face and Gesture Recognition, 196-301, Zürich, 1995.
3. Hashimoto K., ed. *Visual servoing*, World Scientific Publ. Co. Pte. Ltd., London, 1993.



4. Hendzel J. K., *Słownik Polskiego Języka Miganego*, Wydawnictwo OFFER, Olsztyn, 1995.
5. Luong Q.-T., *Color in computer vision* [w:] Chen C. H., Pau L. F., Wang P. S. P., eds, *Handbook of Pattern Recognition and Computer Vision*, 311-368, World Scientific Publ. Co. Pte. Ltd., London, 1993.
6. Marnik J., *Rozpoznawanie znaków Polskiego Alfabetu Palcowego z wykorzystaniem morfologii matematycznej i sieci neuronowych*, rozprawa doktorska, Akademia Górniczo-Hutnicza, Kraków, 2002.
7. Marnik J., *Rozpoznawanie znaków Polskiego Alfabetu Palcowego z wykorzystaniem morfologii matematycznej*, *Archiwum Informatyki Teoretycznej i Stosowanej*, 11: 157-169, 1999.
8. Nieniewski M., *Morfologia matematyczna w przetwarzaniu obrazów*, Akademicka Oficyna Wydawnicza, Warszawa, 1998.
9. Sangwine S. J., Horne R. E. N., *The colour Image Processing Handbook*, Chapman & Hall, London, 1998.
10. Sonka M., Hlavac V., Boyle R., *Image Processing Analysis and Machine Vision*, Chapman & Hall, London, 1994.
11. StatSoft, *STATISTICA Neural Networks*, StatSoft Inc., 1998.
12. Tadeusiewicz R., Korohoda P., *Komputerowa analiza i przetwarzanie obrazów*, Wyd. Fund. Post. Telekom., Kraków, 1997.