

Mirosław FORCZEK
Politechnika Śląska, Instytut Informatyki

ANALIZA WYMAGAŃ DLA PROCESU ROZPOZNAWANIA ZNAKÓW ALFABETU PALCOWEGO Z OBRAZÓW STATYCZNYCH¹

Streszczenie. Przedstawiona została klasyfikacja znaków alfabetu palcowego według własności wizualnych. Znaki zostały podzielone na grupy, w których do rozpoznania symbolu konieczne jest wyznaczenie określonego podzbioru cech. Wybór metod analizy obrazu jest zależny od rodzaju cech koniecznych do rozpoznania symbolu. Wskazane zostały metody, które należy zastosować w systemie automatycznego rozpoznawania.

ANALYZE OF REQUIREMENTS FOR FINGER SPELLING ALPHABET SIGNS RECOGNITION FROM STATIC IMAGES

Summary. The classification of signs from finger spelling alphabet is shown. The signs were classified by their visual properties and separated into groups. In order to recognize sign from a given group, a particular features must be determined. Selection of image analysis methods depends on kind of features required for recognition. Particular methods were mentioned, which should be applied in an automated recognition system.

1. Wprowadzenie do problemu

Alfabet palcowy stanowi podstawowy zbiór symboli języka migowego [1-3]. Z ich pomocą możliwe jest wyrażenie dowolnego pojęcia przez złożenie go z odpowiedniego ciągu znaków – liter. Jakkolwiek w normalnym dialogu za pomocą języka migowego stosowane są

¹ Praca została zrealizowana w ramach projektu badawczego 8T11C 007 17, finansowanego przez Komitet Badań Naukowych w latach 1999-2001.

bardziej zaawansowane znaki – oznaczające całe pojęcia, zamiast pojedynczych liter – opanowanie alfabetu palcowego pozwala już na rozpoczęcie jakiejkolwiek komunikacji. Rozważając konstrukcję systemu rozpoznawania języka migowego, naturalnym podejściem wydaje się rozpoczęcie prac od skończonego zbioru znaków, który będzie umożliwiał podstawy komunikacji – znaków alfabetu palcowego.

1.1. Dane źródłowe

Źródłem danych do niniejszej analizy są dwa zestawy obrazów znaków alfabetu palcowego (dla języka francuskiego): pierwszy zestaw zawiera rysunki, drugi rzeczywiste fotografie [4]. Obrazy przedstawiają znaki alfabetu palcowego dla języka francuskiego, jednak ze względu na duże podobieństwa alfabetów palcowych dla różnych języków wyniki niniejszej analizy powinny być w równym stopniu użyteczne podczas prac nad alfabetem dla języka polskiego.

1.2. Warunki obserwacji

Na potrzeby niniejszej analizy zostały przyjęte następujące założenia dotyczące warunków obserwacji rzeczywistej sceny i akwizycji obrazu wideo:

- dłoń jest jedynym widocznym obiektem,
- tło obrazu jest jednolite,
- tło obrazu jest kontrastowe względem barwy dłoni.

Nie przyjęto żadnych szczególnych założeń co do warunków oświetlenia sceny. Natężenie światła powinno być wystarczające do rozpoznania podstawowych szczegółów widocznych na obrazie. Szczególne rodzaje światła oraz ich położenia przestrzenne nie są tutaj rozpatrywane.

2. Tryby rozpoznawania

Należy brać pod uwagę dwa różne tryby rozpoznawania. Celem procesu rozpoznawania jest dokonanie klasyfikacji danego obrazu jako reprezentacji konkretnego znaku. Wynik klasyfikacji zależy od założonego zbioru możliwych decyzji. Wynikają stąd dwa możliwe tryby pracy zależnie od dziedziny wyniku R :

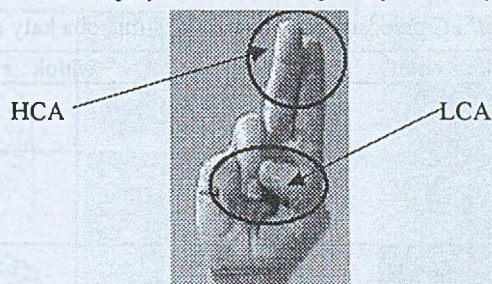
- $R = \{A, \dots, Z\}$, celem tego trybu jest przypisanie obrazowi kodu znaku, który wykazuje największe podobieństwo spośród wszystkich znanych znaków (alfabetu palcowego). Tryb ten zakłada, że obraz przedstawia poprawny znak alfabetu palcowego. Klasyfikator może podjąć decyzje w oparciu o maksymalne podobieństwo.

- $R = \{A, \dots, Z, \Theta\}$, gdzie Θ oznacza znak nie należący do alfabetu palcowego. W tym trybie pracy klasyfikator musi nie tylko określić maksymalne podobieństwo pomiędzy obrazem a konkretnym znakiem, ale musi również zweryfikować podjętą decyzję. Tym razem obraz może zawierać dłoń w przypadkowej (nie znaczącej) konfiguracji.

3. Obszary wysokiego i niskiego kontrastu

Obiekt (lub jego część) może pozostawać w *obszarze wysokiego kontrastu* (HCA - ang. *High Contrast Area*), jeśli jest dobrze widoczny względem tła: obiekt i tło mogą być łatwo odróżnione na podstawie wartości ich atrybutów (np.: kolor, jasność itp.).

Przeciwnym przypadkiem jest pozostawanie obiektu w *obszarze niskiego kontrastu* (LCA – ang. *Low Contrast Area*). Taka sytuacja zdarza się często w przypadku obserwacji dłoni. Poszczególne części dłoni mogą nakładać się na inne (np.: palce). Ponieważ cała dłoń jest tego samego koloru, odróżnienie nakładających się części staje się problematyczne [5,6].



Rys. 1. Obszary wysokiego i niskiego kontrastu na obrazie dłoni

Fig. 1. High and low contrast areas on hand image

Istnienie kontrastu jest podstawowym czynnikiem niezbędnym do jakiegokolwiek przetwarzania lub rozpoznawania obrazu. Wartość kontrastu pomiędzy dwoma elementami (np.: obiekt i tło) musi być większa od poziomu zakłóceń pochodzących z procesu akwizycji obrazu. W przeciwnym przypadku elementy obrazu staną się nierozróżnialne.

Wynikowa wartość kontrastu zależy od wielu czynników: warunków oświetlenia, rozdzielczości kamery, głębi kwantyzacji itd.

Podsumowując, znaki powinny być rozpoznawane za pomocą cech występujących w obszarach wysokiego kontrastu.

4. Cechy wizualne znaków

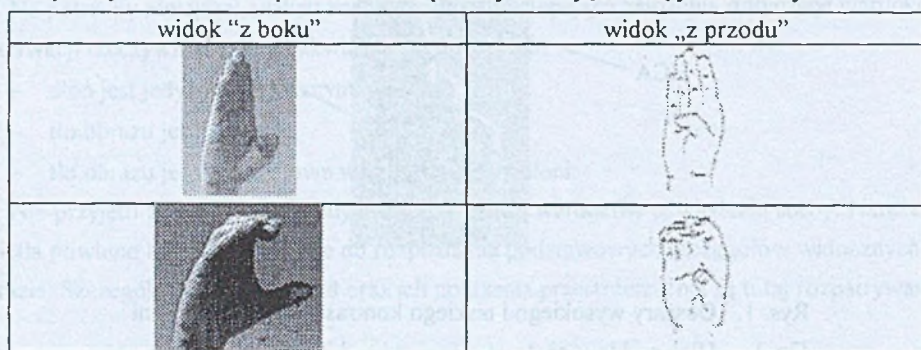
Poniższy rozdział jest próbą określenia cech charakterystycznych [7] dla każdego znaku rozważanego alfabetu palcowego. Znaki zostaną podzielone na grupy według ich cech wizualnych oraz możliwych błędnych klasyfikacji.

4.1. Kąt obserwacji

Mając na uwadze własności obszarów wysokiego i niskiego kontrastu, można zauważyć, iż dany znak będzie posiadał różne cechy wizualne zależnie do kąta obserwacji. Dla celów niniejszej analizy zdefiniowane zostaną dwa główne kąty obserwacji:

- widok „z przodu” (FV – ang. *Front Viewing*), płaszczyzna dłoni jest prostopadła do kierunku obserwacji,
- widok „z boku” (SV – ang. *Side Viewing*), płaszczyzna dłoni jest równoległa do kierunku obserwacji.

Przykłady znaków *B* i *C* przedstawione poniżej ilustrują oba kąty obserwacji:



Rys. 2. Przykłady kątów obserwacji

Fig. 2. Examples of viewing angles

Cechy wizualne obu znaków mogą być łatwo wyróżnione w widoku SV, podczas gdy w widoku FV oba znaki stają się podobne do siebie wzajemnie ze względu na wystąpienie obszarów o niskim kontraście.

4.2. Grupy znaków

4.2.1. Grupa {*A*, *L*, *S*, (*D*, *U*, *G*), *P*, *I*}

Do tej grupy zaliczone zostały znaki, których klasyfikacja powinna być najłatwiejsza. Ich wizualne cechy charakterystyczne leżą w obszarach wysokiego kontrastu.

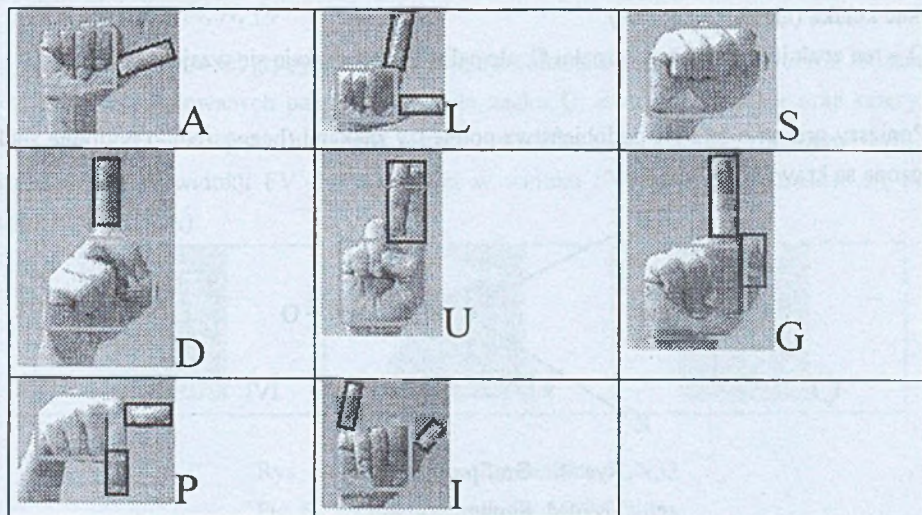
S – ten znak przedstawiany jest przez zamkniętą dłoń. Jego wizualną cechą jest regularny kształt (brak odstających elementów).

A – podobnie jak w przypadku znaku S dłoń jest zamknięta, ale kciuk pozostaje wyprostowany (HCA).

L – podobnie jak dla znaku A, ale również palec wskazujący jest wyprostowany (HCA). Kciuk oraz palec wskazujący są wzajemnie prostopadłe.

D – w tym znaku wyprostowany jest tylko palec wskazujący. Niestety istnieją dwa inne znaki, które mogą być łatwo pomyłone ze znakiem D. Są nimi znaki U oraz G.

U –znak reprezentowany przez wyprostowane dwa sąsiednie palce (wskazujący i środkowy). Można go odróżnić od znaku D przez pomiar szerokości wyprostowanych palców (dwa palce zamiast jednego) lub przez detekcję dwóch czubków palców w tej części obrazu (bardziej skomplikowane od pomiaru szerokości). Należy zauważyć, że oba znaki można wzajemnie rozróżnić w widoku FV. W widoku SV stają się nierozróżnialne.



Rys. 3. Cechy wizualne znaków

Fig. 3. Visual features of signs

G – ten znak jest podobny do znaku D, ale kciuk jest wyprostowany prostopadle do palca wskazującego oraz styka się z nim. Odróżnienie tego znaku od znaku D musi nastąpić przez detekcję czubka kciuka (oprócz czubka palca wskazującego). Detekcja taka może być zawodna – kciuk może zostać zaliczony do obszaru płaszczyzny dłoni lub jego czubek może nakładać się z palcem wskazującym (LCA).

P – ten znak posiada taką samą konfigurację palców jak znak L, ale palec wskazujący skierowany jest do dołu (do góry w przypadku znaku L).

I – ten znak posiada dwa palce wyprostowane: mały oraz kciuk.

Rysunek 3 przedstawia przykłady poszczególnych znaków z oznaczonymi cechami wizualnymi.

4.2.2. Grupa {B, C, E, O}

W tej grupie znajdują się znaki podobne do siebie. Niemożliwe jest określenie ścisłego podziału pomiędzy ich cechami wizualnymi.

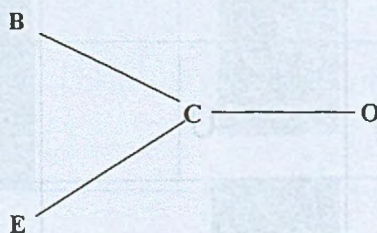
B – wszystkie palce są wyprostowane z wyjątkiem kciuka, który jest zgięty w płaszczyźnie dłoni.

C – cztery palce wraz z kciukiem są lekko zakrzywione. Układ dłoni powinien przypominać literę C w widoku SV.

E – ten znak jest podobny do znaków B oraz C, palce są zakrzywione w podobny sposób jak dla znaku C, ale znajdują się znacznie bliżej płaszczyzny dłoni. Czubki palców mogą dotykać kciuka (ale niekoniecznie).

O – ten znak jest podobny do znaku C, ale palce i kciuk stykają się wzajemnie.

Poniższy graf przedstawia podobieństwa pomiędzy znakami (bezpośrednio podobne znaki połączone są krawędzią).



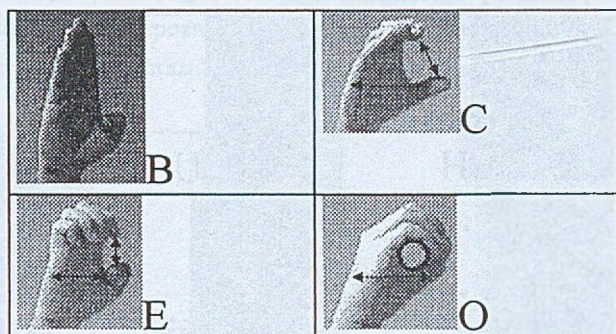
Rys. 4. Graf podobieństw

Fig. 4. Similarity graph

Znaki te powinny być rozpoznawane z wykorzystaniem widoku SV, gdzie stopień zakrzywienia palców jest najlepiej widoczny. Poszczególne znaki różnią się wartościami dwóch parametrów:

- stopniem ugięcia palców: od zupełnie wyprostowanych (B), poprzez częściowo ugięte (C,E) do pełnego ugięcia i styku z kciukiem (O),
- odległością palców od płaszczyzny dłoni: od palców położonych w płaszczyźnie dłoni (B,E) do palców wysuniętych poza płaszczyznę dłoni (C,O).

Nie istnieją ściśle granice zakresów wartości ww. parametrów, które pasują do konkretnego znaku. Przeciwnie, granice pomiędzy znakami mają charakter rozmyty.



Rys. 5. Rozmyte granice pomiędzy znakami

Fig. 5. Fuzzy boundaries between signs

4.2.3. Grupa {M, N, Q}

Wszystkie znaki z tej grupy charakteryzują się palcami skierowanymi do dołu. Różnią się one liczbą wyprostowanych palców: jeden dla znaku Q, dwa dla znaku N oraz cztery dla znaku M. Ponieważ jest to jedyna cecha wizualna tych znaków, powinny być one rozpoznawane w widoku FV. Jednak nawet w widoku FV palce mogą znaleźć się przed nadgarstkiem (LCA).



Rys. 6. Cechy wizualne znaków M, N, Q

Fig. 6. Visual features for M, N, Q signs

4.2.4. Grupa {H, U, V, W}

Znaki z tej grupy reprezentowane są za pomocą dwóch do trzech palców wyprostowanych i zorientowanych w określonej konfiguracji.

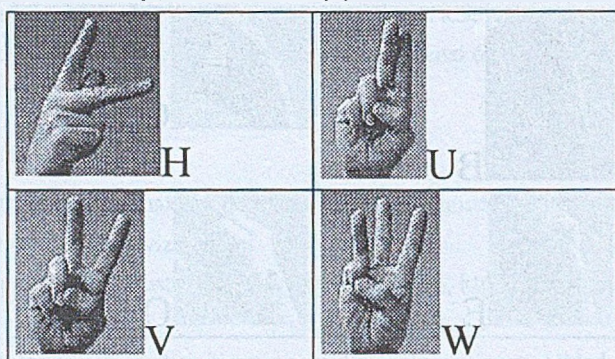
U – dwa sąsiednie palce są wyprostowane (wskazujący i środkowy) oraz stykają się wzajemnie.

V – podobnie jak dla znaku U, ale palce są rozsunięte.

H – podobnie jak dla znaku U, ale palec środkowy znajduje się w płaszczyźnie poziomej.

W – podobnie jak dla znaku U, oraz dodatkowo wyprostowany serdeczny palec.

Znaki te powinny być rozpoznawane w widoku FV. Należy również zauważyć, że granice pomiędzy znakami U, V i H mają charakter rozmyty.

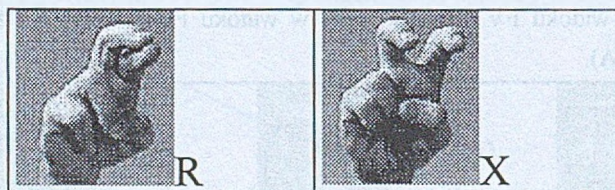


Rys. 7. Znaki H, U, V, W

Fig. 7. H, U, V, W signs

4.2.5. Grupa {R, X}

Oba znaki reprezentowane są przez dwa palce: wskazujący oraz środkowy. Pozostałe palce są zamknięte w płaszczyźnie dłoni. Znaki różnią się jedynie odległością pomiędzy wyprostowanymi palcami:



Rys. 8. Znaki R, X

Fig. 8. R, X signs

- palce są złączone dla znaku R,
- palce są rozsunięte dla znaku X.

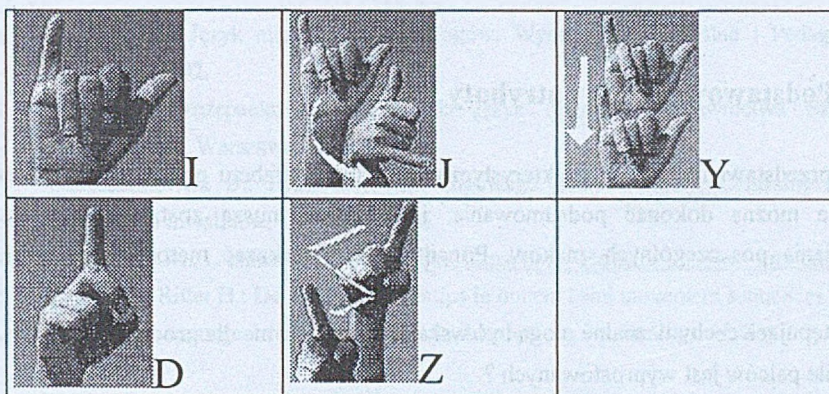
Ponownie, granica pomiędzy znakami (palce złączone – palce rozsunięte) ma charakter rozmyty.

4.2.6. Grupy {I, (J, Y)} i {D, (Z)}

Grupy te zawierają znaki dynamiczne. Każda grupa ma swój statyczny prymityw: znaki I oraz D, które zostały już przypisane do pierwszej z omawianych w tym rozdziale grup. Wprowadzone z nich znaki mają takie same konfiguracje palców oraz orientację płaszczyzny

dłoni jak ich statyczne odpowiedniki. Charakterystycznymi cechami nowych znaków są iteracyjne ruchy całej dłoni według określonej trajektorii.

Znaki te muszą być rozpoznawane w dziedzinie czasu, nie tylko z wykorzystaniem pojedynczego statycznego obrazu. Z tego względu wykraczają one poza obszar niniejszej analizy.



Rys. 9. Znaki dynamiczne

Fig. 9. Dynamic signs

4.2.7. Grupa {F, K, T}

Ostatnia z grup zawiera znaki najtrudniejsze do rozpoznania. Konfiguracje palców tych znaków zawierają kombinacje cech wizualnych, które należy obserwować w widokach FV i SV.



Rys. 10. Znaki F, K, T

Fig. 10. F, K, T signs

F – trzy palce są wyprostowane i skierowane do góry (FV), palec wskazujący jest wyprostowany i zlokalizowany w płaszczyźnie poziomej, kciuk jest wyprostowany i dotyka palca wskazującego (SV).

K – dwa palce (wskazujący i środkowy) są wyprostowane i rozsunięte (FV). Kciuk jest również wyprostowany do góry i umiejscowiony pomiędzy palcami wskazującym i środkowym (FV), przed płaszczyzną dłoni (SV).

T – trzy palce są wyprostowane do góry (FV), ponadto palce wskazujący i kciuk stykają się czubkami (SV).

5. Podstawowe cechy i atrybuty wizualne

Po przedstawieniu cech charakterystycznych znaków alfabetu palcowego w poprzednim rozdziale można dokonać podsumowania: jakie cechy muszą zostać wykorzystane do rozróżniania poszczególnych znaków. Ponadto należy wskazać metody ich pozyskania z obrazu.

Następujące cechy wizualne mogą być wskazane jako istotne dla procesu rozpoznawania:

- ile palców jest wyprostowanych ?
- które palce są wyprostowane ?
- jaki jest stopień ugięcia palców ?
- jaka jest orientacja dłoni ?
- czy palce są skierowane do góry czy do dołu ?
- jaka jest pozycja kciuka ?
- czy kciuk dotyka innych palców ?

Wyżej wymienione cechy są silnie związane z fizyczną budową ludzkiej dłoni. Aby móc odpowiedzieć na tak postawione pytania, należy dokonać detekcji cech podstawowych z płaskiego obrazu. Następnie, wykorzystując wiedzę o budowie i mechanice dłoni, można wnioskować o bieżącej konfiguracji dłoni, która najbardziej pasuje do obserwowanych cech podstawowych [8].

Następujące cechy można zaproponować jako podstawowe dla obrazu płaskiego dłoni:

- położenia czubków palców, jeśli zostały odnalezione,
- położenia segmentów palców, jeśli zostały odnalezione,
- obszar dłoni: kształt i rozmiar.

Powyższe cechy nadal pozostają na wysokim poziomie abstrakcji względem typowych metod przetwarzania obrazu [9,10]. Kolejnym krokiem jest więc wskazanie środków służących do ekstrakcji tych cech z obrazu.

Można zaproponować wykorzystanie trzech rodzajów źródeł danych:

- kolor – dla celów segmentacji [11],

- detekcja krawędzi – dla celów segmentacji oraz detekcji konturów [12],
- analiza konturów – dla celów lokalizacji segmentów i czubków palców [13-18].

LITERATURA

1. Pietrzak W.: Język migowy dla pedagogów. Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1992.
2. Perlin J., Szczepankowski B.: Polski język migowy. Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1983.
3. Szczepankowski B.: Lektorat języka migowego. Kurs wstępny. Centralny Związek Spółdzielni Inwalidów, Warszawa 1986.
4. Vannobel J.M.: L'alphabet dactylogique, <http://www-i3d.univ-lille1.fr/~jmv/dactylo/>
5. Nolker C., Ritter H.: Detection of fingertips in human hand movement sequences. *Gesture and Sign Language in Human-Computer Interaction, Proceedings of GW'97*, Springer 1997.
6. Triesch J., Malsburg Ch.: Robust Classification of Hand Postures against Complex Backgrounds. *IEEE Computer Society Press, Proceedings of the Second International Conference on Automatic Face and Gesture Recognition '96*, Killington, Vermont 1996.
7. Losson O., Vannobel J.M.: Sign Specification and Synthesis. *Gesture-Based Communication in Human-Computer Interaction, Proceedings of GW'99*, Springer 1999.
8. Nolker C., Ritter H.: GRAFIT: visual recognition of hand postures. *Gesture-Based Communication in Human-Computer Interaction, Proceedings of GW'99*, Springer 1999.
9. Watkins Ch., Sadun A., Marenka S.: Nowoczesne metody przetwarzania obrazów. Wydawnictwa Naukowo-Techniczne, Warszawa 1995.
10. Tadeusiewicz R, Korohoda P.: Komputerowa analiza i przetwarzanie obrazów. Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1997.
11. Chen C., Chiang S.: Detection of human faces in colour images. *IEE Proceedings – Vision, Image and Signal Processing*, t. 144, nr 6, 1997.
12. Antoniu E., Sandu I.: On pros and cons of implementing '80s theory of edge detection. *Machine GRAPHICS and VISION*, t. 1, nr 1-2, s. 255-260, 1992.
13. Kierzenkowski K., Puchalski J.: Object contour recognition using oriented grids. *Machine GRAPHICS and VISION*, t. 3, nr 4, s. 657-665, 1994.
14. Zhu Y., Seneviratne L.: Optimal polygonal approximation of digitized curves. *IEE Proceedings – Vision, Image and Signal Processing*, t. 3, nr 4, 1997.
15. Ip H., Wong W.: Fast conditioning algorithm for significant zero curvature detection. *IEE Proceedings – Vision, Image and Signal Processing*, t. 144, nr 1, 1997.

16. Wilson G.: Properties of contour codes. IEE Proceedings – Vision, Image and Signal Processing, t. 144, nr 3, 1997.
17. Fu A.M.N, Yan H.: Effective classification of planar shapes based on curve segment properties. Pattern Recognition Letters, t. 18, 1997
18. Boomgaard R.: Affine invariant deformation curves a tool for shape characterization. Image and Vision Computing, t. 18, 1999.

Recenzent: Dr inż. Henryk Palus

Wpłynęło do Redakcji 20 czerwca 2001 r.

Abstract

The finger spelling alphabet is the base symbol set for deaf language. When we consider construction of a system for deaf language recognition, it is natural to start the work from finite symbols set, which allow for basic communication. As source for this analysis two image collections were taken: first collection contains drawings and second contain real photographs.

We need to consider two recognition modes: $R = \{A, \dots, Z\}$ - the target of this mode is to assign to a given image code of symbol, for which the image corresponds with maximum similarity. The assumption is, the image shows a valid symbol of the finger spelling alphabet. With other mode: $R = \{A, \dots, Z, \Theta\}$, where Θ means symbol that doesn't belong to the finger spelling alphabet, the classifier must not only determine the maximum similarity between an image and one of known symbols, but also has to verify its decision.

An object (or its part) may stay in **high contrast area**, if it is clearly visible against background. Otherwise, an object stays in **low contrast area**. Such case often exists by observation of human hand: the parts of the hand can overlap other ones (e.g. fingers). Because all the hand has the same color, it is very difficult to distinguish between overlapped parts. An existence of contrast is essential for any image processing or recognition. The value of contrast between two elements (object and background) must be higher then noise level that comes from image acquisition. Otherwise, two elements will be undistinguishable.

It can be seen, that certain sign will have different visual features, depending on viewing angle. For purposes of this analyze, we define two main viewing angles: the **front viewing (FV)** when the palm plane is perpendicular to viewing direction, and the **side viewing (SV)** when the palm plane is parallel to viewing direction.