

LOKALIZACJA KLUCZOWYCH CECH ROZPOZNAWANYCH OBIEKTÓW NA PODSTAWIE ANALIZY ŚCIEŻKI WZROKOWEJ OBSERWATORA

LOCALIZATION OF THE OBJECT'S SALIENT FEATURES BASED ON VISUAL PATH ANALYSIS

ZBIGNIEW MIKRUT, PIOTR AUGUSTYNIAK

Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, zibi@agh.edu.pl, august@agh.edu.pl

Streszczenie. W artykule przedstawiono metodykę badań, których celem jest wyznaczenie istotnych (z punktu widzenia rozpoznawania) cech obiektów. Odpowiednie eksperymenty przeprowadzono przy pomocy aparatury OBER2, służącej do rejestracji ruchu oczu człowieka. Zaplanowano i zrealizowano eksperymenty rozpoznawania ręcznie pisanych cyfr oraz opisano metodę lokalizacji cech na podstawie dwóch wskaźników.

Słowa kluczowe: cechy obiektów, lokalizacja cech, ścieżka wzrokowa, sygnał okoruchowy

Abstract. In the present paper the methodology of the salient features detection is presented. The experiments were executed using the OBER2 eye-tracker. During the particular experiment students to be examined had to recognize the more or less blurred images of handwritten digits. The salient features localization method is based on two factors derived from the visual path analysis.

Key words: salient features, feature localization, eye tracking, visual path analysis

1. Wstęp

Modelowanie fragmentów systemu wzrokowego człowieka i zwierząt wyższych może być realizowane na różnych poziomach szczegółowości, począwszy od modelu rozmieszczenia receptorów i pól receptorowych (transformacja log-polar [5]) poprzez analizę przepływu informacji traktem wzrokowym [7] na ogólnych aspektach rozumienia obserwowanej sceny skończywszy [6]. Celem niniejszej pracy jest zbadanie, czy jest możliwa lokalizacja cech obiektów, na jakie zwraca uwagę człowiek podczas procesu rozpoznawania, na podstawie analizy ścieżek wzrokowych jego oczu.

Do badań wykorzystano urządzenie OBER2, skonstruowane na Politechnice Śląskiej w Gliwicach [4]. Badaniem zakłada się rodzaj okularów bez szkieł („gogle”), na których są umieszczone dwie pary oświetlaczy podczerwieni oraz tyle samo odbiorników odbitego promieniowania. Diody oświetlają gałkę oczną, a odbiorniki rejestrują odbite promieniowanie. Odbywa się to niezależnie w dwóch wzajemnie prostopadłych osiach. Ruch gałki ocznej powoduje zmiany natężenia odbitego promieniowania.

Sygnały z „gogli” dochodzą do przetwornika analogowo-cyfrowego, następnie są zapisywane w buforze urządzenia i transmitowane łączem szeregowym do komputera klasy PC. Aparatura OBER2 dokonuje akwizycji z maksymalną

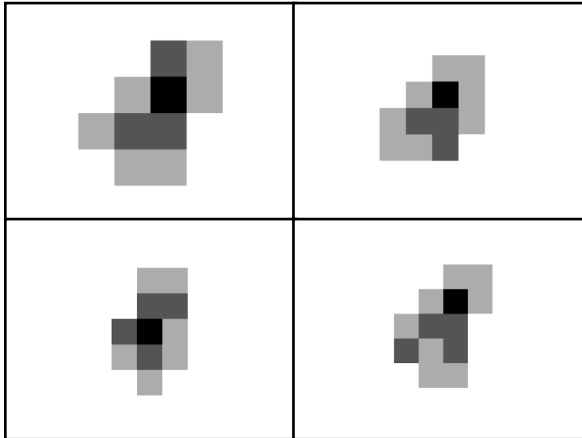
częstotliwością próbkowania 2 kHz przy 12 bitowej rozdzielczości kwantowania.

Konstruktorzy aparatury OBER2 dostarczają wraz z urządzeniem podstawowe oprogramowanie pracujące pod systemem DOS i umożliwiające dynamiczną kalibrację oraz akwizycję ścieżki wzrokowej podczas prezentacji statycznych obrazów. W Laboratorium Biocybernetyki AGH w Krakowie powstała rozszerzona wersja oprogramowania, działająca w systemach Windows. Umożliwia ona kalibrację statyczną (badany ma za zadanie obwieść wzrokiem wyświetlony na monitorze prostokąt) oraz wielokrotną akwizycję ścieżki wzrokowej w trakcie ściśle zaplanowanego eksperymentu. Pojedynczy eksperyment może składać się z pokazu filmu video lub prezentacji sekwencji różnych obrazów, przy czym zarówno kolejność, jak i czasy ekspozycji są z góry zadane w odpowiednim pliku konfiguracyjnym. Tę ostatnią możliwość wykorzystano w opisywanych badaniach.

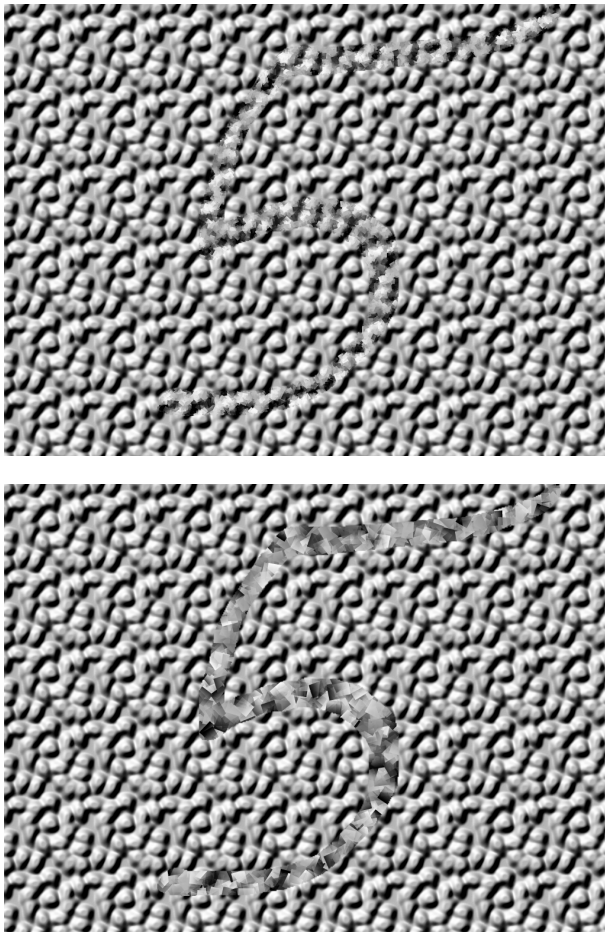
2. Eksperymenty

Cel pracy, czyli badanie możliwości lokalizacji istotnych cech obiektów, był realizowany na przykładzie rozpoznawania ręcznie pisanych cyfr. Obydwa te problemy (lokalizacji cech i rozpoznawania) są intensywnie badane, lecz nie do końca rozwiązane [1].

W celu uniknięcia sytuacji rozpoznawania znaku „jednym rzutem oka” zaplanowano dwie serie eksperymentów, różniące się rodzajem prezentowanych sekwencji obrazów. Prezentacje polegały na kolejnych pokazach obrazów o coraz większej wierności odwzorowania. W serii pierwszej zastosowano tzw. pikselizację, czyli uśrednianie całego obrazu znaku (czarny znak na białym tle) w oknach o zmiennej wielkości. Na rys. 1 przedstawiono kilka kolejnych obrazów cyfry „1”. W serii drugiej znaki „wylaniały się” z obrazu tekstury, czyli były rysowane przy pomocy tekstur coraz bardziej różniących się od tła (por. rys. 2).



Rys. 1. Przykład serii obrazów (cyfra „1”) o malejącym stopniu trudności rozpoznawania otrzymanych metodą „pikselizacji”



Rys. 2. Dwa obrazy (cyfra „5”) o malejącym stopniu trudności rozpoznawania otrzymane metodą „teksturową”.

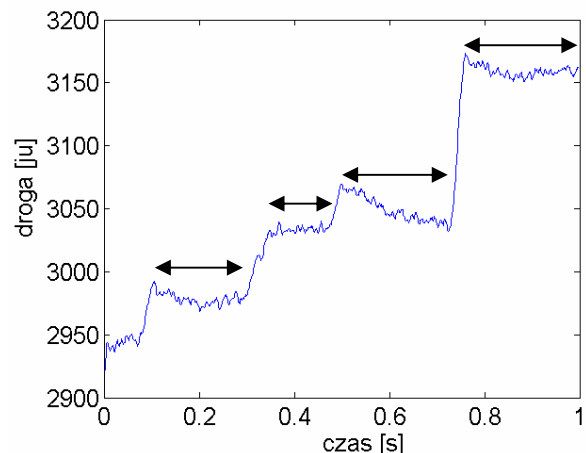
Eksperymenty polegały na przeprowadzeniu kalibracji statycznej (por. pkt 1), sprawdzeniu jej wyników, a następnie na rejestracji ruchów oczu podczas pokazu kolejnych obrazów danego znaku. Czasy ekspozycji kolejnych obrazów zmieniły się od około 5 s dla projekcji „najtrudniejszych” do ok. 2 s dla obrazów mało zakłóconych. Urządzenie OBER2 dokonywało akwizycji poszczególnych składowych ścieżki wzrokowej z częstotliwością 750 Hz.

3. Analiza wyników

W poprzednich pracach autorów głównym kryterium zainteresowania badanego był czas koncentracji uwagi na danym fragmencie obrazu [2], [3]. Podczas analizy sygnałów ruchu oczu, przeprowadzonej w ramach niniejszej pracy, zdecydowano się (głównie z przyczyn wizualizacyjnych) na zastosowanie tego kryterium w formie wydzielenia „obszaru zainteresowania”, czyli wyodrębnienia fragmentów sceny, na których oko obserwatora (przez pewien czas i w sposób ciągły) *najdłużej* wykonywało tylko ruchy z małą prędkością.

Postać drugiego kryterium wyniknęła z metody prowadzenia eksperymentów. Podczas przeglądania wyników okazało się, że badani – nie będąc pewni swojego rozpoznania – wielokrotnie powracali wzrokiem do „niepewnych” fragmentów obrazu cyfry. Można postawić hipotezę, że cechy te stanowią elementy różnicujące poszczególne cyfry. Tak więc jako drugie kryterium przyjęto „liczbę powrotów” uwagi badanego w dany rejon sceny. W ten sposób wyznaczono drugi obszar sceny o szczególnej zawartości informacyjnej – obszar, dla którego „liczba powrotów” była największa.

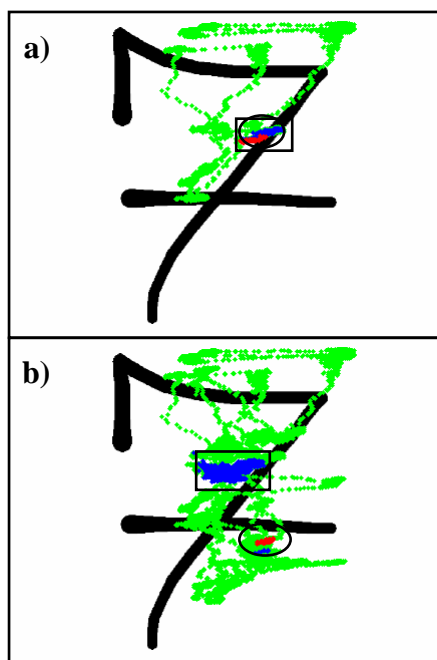
Algorytm obliczania obu wskaźników opiera się na analizie chwilowej prędkości ruchu oka, wspomaganą informacją o jego aktualnym położeniu. Przykładowy przebieg drogi chwilowego punktu koncentracji uwagi jednego oka pokazano na rys. 3. Jak widać, można względnie łatwo wyodrębnić odcinki ruchu wolnego, odpowiadającego obserwacji danego fragmentu sceny (na rys. 3 fragmenty te zaznaczono strzałkami) oraz szybkiego, oznaczającego przenoszenie punktu koncentracji uwagi w inny obszar.



Rys. 3. Fragment rejestracji drogi punktu koncentracji uwagi (oko prawe). [ju] oznacza „jednostki umowne” współrzędnych obserwowanej sceny

Punkty odpowiadające kolejnym odcinkom ruchu wolnego tworzą kolejne obszary skupienia punktu koncentracji uwagi. Po zakończeniu wyszukiwania, określone są granice każdego

z obszarów, a następnie – dla każdego znalezionej obszaru – sprawdzane jest, ile razy dany obszar nałożył się na inny. W końcu wybierany jest obszar największy oraz obszar o największej liczbie nałożeń, czyli powrotów punktu koncentracji uwagi. Obszary te to wspomniane wyżej wskaźniki lokalizacji cech rozpoznawanego obiektu. Zostają one naniesione na źródłowy obraz cyfry. Przykład lokalizacji cech pokazano na rys. 4.



Rys. 4. Wizualizacja dwóch wskaźników lokalizacji cech: obszaru najdłuższego skupienia uwagi (oznaczony elipsą) oraz obszaru o największej liczbie powrotów „uwagi” obserwatora (obszar oznaczony prostokątem): a) do momentu rozpoznania (1 obraz), b) dla całego eksperymentu. Obraz b) został uzyskany w wyniku analizy rejestracji osoby 5 (por. tab. 1).

Na podstawie wizualizacji możliwe jest określenie kluczowych cech, na jakie zwracał uwagę badany. W tabeli 1 zestawiono przykładowe wyniki uzyskane od 10 badanych osób dla rozpoznawania cyfr „1” i „7”. Wyniki te otrzymano, analizując cały eksperyment, włącznie z trajektoriami zarejestrowanymi po rozpoznaniu obiektu (jak na rys. 4b). Wybrano akurat tę parę cyfr, ponieważ są one dość podobne – szczególnie na początkowych etapach prezentacji. Obszar obrazu cyfry podzielono na 4 podobszary oznaczone kolejnymi literami alfabetu (por. rys. 5):

a – środek obiektu, w przypadku cyfry „7” odpowiada poziomej kresce,

b – górna część cyfry,

c – część lewa (okolice daszka),

d – podobszar poza obrysem.

Pomimo niewielkiej (statystycznie) liczby badanych i eksperymentów można pokusić się o wstępną interpretację zestawionych wyników.

W ponad połowie eksperymentów badani wybierali takie same cechy dla obu wskaźników, czyli obszar skupienia wypadł w pobliżu obszaru częstych powrotów punktu koncentracji uwagi. W tej grupie dwie osoby (nr 9 i 10) szukały cech z użyciem obu metod w tym samym obszarze

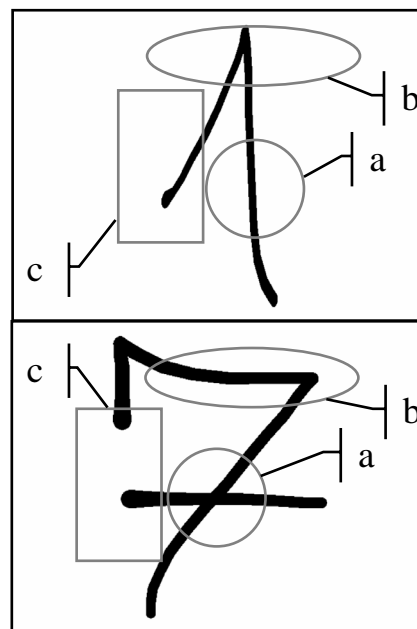
w przypadku obu cyfr. Był to obszar „b”, czyli górna część znaku. W przypadku cyfr „1” i „7” odpowiada to nachyleniu „daszka”, co stanowi bardzo ważną cechę różnicującą także w przypadku innych obserwatorów.

Na obszarze „c”, czyli na lewej części obrazu znaku, skoncentrowano uwagę podczas 9 eksperymentów. W 4 przypadkach obszary najdłuższej koncentracji uwagi i największej liczby powrotów pokrywały się.

Kolejną cechą jest pozioma kreska na obrazie cyfry „7” (obszar „a”). Rejon ten był przeszukiwany podczas 5 eksperymentów (na 20 przeprowadzonych), przy czym w 3 przypadkach zachodziła sytuacja pokrywania się obszarów określonych obydwoma wskaźnikami.

Tab. 1. Zestawienie położenia kluczowych cech cyfr „1” i „7” dla 10 badanych osób (objaśnienia w tekście i na rys. 5)

lp.	cyfra „1”		cyfra „7”	
	skupienia	„powroty”	skupienia	„powroty”
1.	b	c	c	c
2.	a	a	b	b
3.	c	c	d	c
4.	c	c	b	b
5.	a	b	a	a
6.	a	c	b	c
7.	d	d	c	c
8.	a	a	c	b
9.	b	b	b	b
10.	b	b	b	b



Rys. 5. Położenie podobszarów, dla których zlokalizowano kluczowe cechy (por. tab. 1): okrąg oznacza podobszar „a”, elipsa – „b”, prostokąt – „c”. Podobszar „d” znajduje się poza obrysem znaku

4. Podsumowanie

Na podstawie przeprowadzonych badań składających się z serii eksperymentów z użyciem urządzenia OBER2 oraz z algorytmów interpretacji otrzymanych rejestracji można stwierdzić, że postawiony na wstępie cel został osiągnięty: okazało się, że jest możliwa lokalizacja cech obiektów, na jakie zwraca uwagę człowiek podczas procesu rozpoznawania

na podstawie analizy trajektorii oczu. Zaproponowane wskaźniki: obszar najdłuższego skupienia wzroku oraz obszar o największej liczbie powrotów uwagi dobrze lokalizują istotne cechy rozpoznawanych cyfr.

Odrębnym zagadnieniem jest interpretacja uzyskanych wyników, których przykład zawiera tabela 1. Na ich podstawie nie można wskazać jednej metody rozpoznawania. Obserwuje się zindywidualizowane strategie rozpoznawania, choć w części przypadków obydwie wskaźniki (obszary) pokrywają się lub znajdują się blisko siebie. Analiza położenia cech dla pojedynczych cyfr lub dla innych par (jak pokazano to w tabeli 1) będzie tematem dalszych badań. Kolejnym etapem interpretacji wyników może być także określenie kolejności przenoszenia uwagi, uwzględnienie większej liczby obszarów (nie tylko największych) oraz badanie otrzymanych rejestracji przedziałami odpowiadającymi np. okresom przed i po rozpoznaniu obiektu.

Podziękowania

Autorzy dziękują studentom III roku kierunku Automatyka i Robotyka Bartłomiejowi Kluzie i Adrianowi Korczyńskiemu za pomoc w realizacji eksperymentów oraz w interpretacji wyników.

Opisane prace zostały zrealizowane w ramach badań statutowych (umowa AGH nr 11.11.120.249).

Literatura cytowana

1. Alpaydin E., Salah A. A., Akarun L.: A Selective Attention Based Method for Visual Pattern Recognition with Application to Handwritten Digit Recognition and Face Recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(3): 420-425, 2002.
2. Augustyniak P., Mikrut Z.: Correlating the degree of observer's preoccupation and the observation time: visual tasks with OBER2 eyetracker, *J. of Medical Informatics & Technologies*, 4: MT3-MT10, 2002.
3. Mikrut Z., Augustyniak P.: Estimation of execution times for tasks of objects counting and localization using the OBER2 device, *Proc. 12th Nordic Baltic Conf. on Biomedical Engineering and Medical Physics*, 144-145, Reykjavik 2002.
4. Ober J., Hajda J., Loska J., Jamicki M.: Application of eye movement measuring system OBER2 to medicine and technology, *Proc. of SPIE, Infrared technology and applications*, , vol. 3061, part 1, pp. 327-33, Orlando, USA 1997.
5. Schwartz E. L.: Spatial mapping in the primate sensory projection: analytic structure and relevance to perception, *Biological Cybernetics*, 25:181-194, 1977.
6. Tadeusiewicz R., Ogiela M. R.: *Medical Image Understanding Technology*, Springer, Berlin 2004.
7. Zeki S.: *Obrazy wzrokowe w mózgu i umyśle*, Świat Nauki, 11:43-51, Warszawa 1992.

