

Lokalizacja istotnych cech obrazu metodą śledzenia ścieżki wzrokowej

Piotr Augustyniak, Zbigniew Mikrut

Katedra Automatyki AGH, Kraków

30-059 Kraków, Mickiewicza 30, {august | zibi}@agh.edu.pl

Streszczenie: Modele perceptualistyczne różnych klas scen znajdują ostatnio uznanie jako wartościowe narzędzie ułatwiające interakcję człowieka z otoczeniem zaawansowanym technologicznie. Przewidywany sposób postrzegania sceny przez człowieka ma istotny wpływ na ergonomiczne projektowanie nowych urządzeń oraz sposobów komunikacji wizualnej. Analiza ścieżki wzrokowej w kontekście informacji wizualnej prowadzi do wyodrębnienia stref sceny zawierających bardziej istotne informacje, a dzięki generalizacji, do sformułowania relacji pozwalającej na oszacowanie i ilościową analizę strumienia informacji pobieranej przez system wizyjny człowieka. Postępowanie takie jest szczególnie przydatne w zastosowaniach wymagających badania oczekiwań obserwatora w sposób niezależny od jego świadomości. Może być również przydatny do badania doświadczenia obserwatora poprzez analizę nawyków okoruchowych i nauczanie interpretacji obrazów drogą niewerbalną. W artykule przedstawiono rezultaty doświadczeń wizualnych przeprowadzonych w Laboratorium Biocybernetyki AGH, których celem była identyfikacja zjawisk zachodzących w systemie okoruchowym towarzyszących postrzeganiu i ocena ich wpływu na związek czasu fiksacji wzroku z ilością informacji pobranej ze sceny. Ostatni z eksperymentów wykazuje proporcjonalną zależność strumienia informacji od czasu fiksacji wzroku w ograniczonym zakresie czasów, co uprawnia do stosowania gęstości fiksacji jako estymatora lokalnej istotności informacyjnej obrazu.

Summary: The perceptual models (PM) of various scenes have been recently recognised as a valuable tool for improving interaction of human with sophisticated devices. The expected manner of scene perception by the human has a remarkable impact for the ergonomic design of new devices and the technologies of visual communication. The analysis of eyeglobe trajectory (scanpath) reveals zones in the visual surrounding that contain more relevant information. Thanks to generalisation, the relation can be established for the estimation and quantitative assessment of data stream acquired by the human visual system. This approach is particularly useful for investigations of the observer's preferences without involvement of his consciousness. Another area of application is the assessment of the observer's skills through the analysis of his oculomotoric habits and the non-verbal teaching of image interpretation. In the paper the visual experiments performed in the Biocybernetics Laboratory AGH-UST and their results are presented. Our goal was the identification of principal phenomena in the oculomotoric system accompanying the perception and the assessment of their influence to the relation of fixation time and the amount of information gathered from the scene. The final experiment proves the proportional dependence of the data stream and the fixation time in a limited time range, that justifies the use of fixation density as the estimate of local relevance of the image.

1. Wprowadzenie

Analiza trajektorii ruchu gałki ocznej podczas wykonywania określonego zadania wizualnego dostarcza wielu informacji o zdolnościach, nawykach i preferencjach obserwatora w poszukiwaniu wiedzy na obrazie [5]. W zależności od sformułowania zadania wizualnego otrzymana ocena może być obiektywna tj. pozbawiona wpływu świadomości obserwatora, a ponadto nie jest zniekształcona niedokładnością werbalizacji. Przybliżenie (modelowanie) sposobu interpretacji obrazu przez eksperta wymagałoby analizy jego procesów myślowych i w chwili obecnej trudno wyobrazić sobie realizację takiego przedsięwzięcia. Zamiast tego wykorzystać można znaną w ergonomii technikę śledzenia ścieżki wzrokowej w celu

identyfikacji nawyków okulomotorycznych wykształconych u eksperta w toku nabywania praktyki [4].

Celem analizy ścieżki wzrokowej obserwatora (ang. *scanpath*) do identyfikacji regionów zainteresowań podczas wykonywania określonego zadania jest określenie ilości informacji pobieranej z poszczególnych stref postrzeganej sceny. Niestety, w literaturze dostępne są jednak tylko jakościowe wnioski dotyczące atrakcyjności prezentowanych elementów określonej związanej z ich barwą, kształtem, położeniem, sąsiedztwem itp. Przeprowadzenie ścisłych pomiarów ilościowych jest bardzo trudne, wymaga bowiem zapewnienia, aby grupa obserwatorów była reprezentatywna dla całej opisywanej populacji m. in. w zakresie własności systemu wizyjnego i okulomotorycznego. Byłoby to związane z przeprowadzeniem u nich kosztownych badań medycznych w zadanych warunkach psychofizycznych.

Określenie ilościowej zależności ilości informacji pobranej przez system wzrokowy człowieka od czasu patrzenia wymaga uwzględnienia zjawisk towarzyszących, zachodzących w systemie okoruchowym ale nie związanych bezpośrednio z percepcją. Wyodrębniono trzy grupy takich zagadnień [2]:

- ścisłe określenie początku i końca obserwacji z uwzględnieniem czasu reakcji obserwatora,
- niejednoznaczność percepcji dwuocnej,
- parametry dynamiczne systemu okoruchowego i dokładność pozycjonowania osi optycznej.

Ponieważ wpływ wymienionych zjawisk jest uwarunkowany cechami osobniczymi obserwatora oraz jego aktualną kondycją psychofizyczną, konieczne jest każdorazowe określenie parametrów korygujących analizę ścieżki wzrokowej. Analiza ta powinna być przeprowadzona automatycznie i bez konieczności wykonania dodatkowych zadań wizualnych.

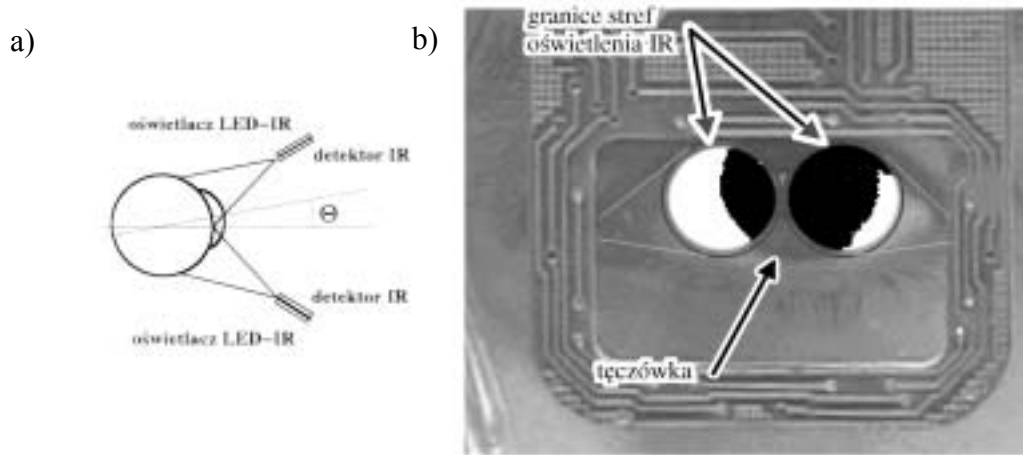
Niniejszy artykuł przedstawia sposób identyfikacji zjawisk towarzyszących percepcji i określenia ich wpływu na przebieg ścieżki wzrokowej. Identyfikacja ta została zrealizowana za pomocą analizy sygnału zarejestrowanego podczas kalibracji urządzenia okoruchowego i może być podstawą eliminacji wpływu tych zjawisk na analizę przemieszczenia punktu uwagi obserwatora. Zasadniczym celem jest opis eksperymentu, w wyniku którego określono zależność ilości informacji pozyskanej ze sceny od czasu prezentacji oraz ograniczenia tej zależności. Na zakończenie przeprowadzono dyskusję otrzymanych rezultatów i podano przykłady zastosowań ścieżki wzrokowej do opisu sposobu postrzegania elementów wybranych klas scen.

2. Metodologia

2.1. Rejestracja ścieżki wzrokowej

Analiza sposobu postrzegania jest zwykle dokonywana przy pomocy urządzenia rejestrującego ruchy gałki ocznej (ang. *eyetracker*) w trakcie realizacji przygotowanych zadań wizualnych (ang. *visual task*). Spośród nieinwazyjnych metod badania położenia gałki ocznej wybrano metodę porównania wiązki promieniowania podczerwonego odbitego od gałki ocznej zastosowaną w urządzeniu OBER2 (rys. 1) [3]. Metoda ta charakteryzuje się dużą precyzją

pozycjonowania oraz odpornością na zakłócenia optyczne (w postaci dodatkowego promieniowania rejestrowanego przez detektory) a także znaczną niewrażliwością na zjawiska elektrofizjologiczne zachodzące wokół gałki ocznej (aktywność elektryczna mięśni powiek, mięśni okulomotorycznych i źrenicy). Oświetlacze oraz detektory promieniowania podczerwonego umieszczone są w goglach zakładanych przez obserwatora, skąd sygnał elektryczny pobierany jest do systemu OBER2 a po przetworzeniu na postać cyfrową (12 bitów, 500 Hz) dostępny na łączu komunikacji szeregowej w postaci ciągu czwórek liczb będących reprezentacją pozycji centrum każdego z oczu w osi pionowej i poziomej [3]. Poprawna interpretacja danych pochodzących z urządzenia jest możliwa dzięki ustaleniu pozycji obserwatora względem prezentowanej sceny (monitora komputerowego) a także kalibracji urządzenia polegającej na obserwacji prostokąta o standaryzowanych wymiarach każdorazowo przed i po pokazie. Ścieżkę wzrokową zarejestrowaną podczas procesu kalibracji wykorzystano do identyfikacji zjawisk towarzyszących percepcji (p. 2.3).



Rys. 1. Pomiar pozycji gałki ocznej na podstawie różnicy natężenia refleksów w podczerwieni; a) zasada fizyczna; b) fragment urządzenia Ober-2

2.2. Zadania wizualne

Prawie cała aktywność człowieka podczas czuwania wypełniona jest realizacją zadań wizualnych polegających na śledzeniu sceny w celu zdobycia pożądaných informacji. Zadania wizualne przeprowadzane podczas eksperymentów okoruchowych charakteryzują się często konkretnym celem (zadaniem) do wykonania. Istotne jest określenie początkowego zasobu wiedzy obserwatora oraz oczekiwanej wiedzy końcowej (wyniku), co umożliwi określenie ilości informacji pozyskanej ze sceny oraz sposobu jej poszukiwania [1, 4]. Podczas realizacji zadań scena postrzegana przez obserwatora jest zazwyczaj ściśle kontrolowana (lub co najmniej rejestrowana), gdyż trajektoria ruchu gałek ocznych musi być analizowana w kontekście zmian sceny. Zadania wizualne mogą narzucać ograniczenie czasu obserwacji, co mobilizuje obserwatora do identyfikacji najistotniejszych elementów sceny w pierwszej kolejności [3], a zatem umożliwia ich lokalizację na podstawie sygnału okoruchowego.

2.3. Identyfikacja zjawisk towarzyszących postrzeganiu

Pozyskiwanie treści przez system wizyjny człowieka napotyka na liczne ograniczenia ze strony mechanizmów fizjologicznych zaangażowanych w ten proces. W rezultacie istotną część czasu zadania wizualnego jest przeznaczona na poszukiwanie celu i pozycjonowanie gałki ocznej przez system okoruchowy. Określenie zjawisk towarzyszących percepcji i wyznaczenie ich udziału w wypadkowej pozycji gałki ocznej jest podstawą interpretacji sygnału okoruchowego jako wskaźnika przemieszczenia punktu koncentracji uwagi obserwatora. Złożoność oraz osobnicza zmienność omawianych zagadnień, znacząco komplikuje algorytmy identyfikacji zjawisk towarzyszących.

Wyznaczenia *czasu reakcji* na rozpoczęcie prezentacji sceny dokonano poprzez wyświetlenie punktu koncentracji w środku ekranu poprzedzającego właściwą prezentację, a następnie detekcję pierwszego świadomego przemieszczenia uwagi w poszukiwaniu celu w każdym z wymiarów. Detekcja momentu zakończenia wykonywania zadania została zrealizowana przy współpracy obserwatora, który jest proszony o skierowanie wzroku poniżej ekranu po wykonaniu zadania.

Wpływ *ograniczeń mechanicznych i sensorycznych* został wyrażony przy pomocy parametrów liczbowych, umożliwiających korektę rezultatów zadania wizualnego:

- *prędkość poszukiwania celu* (maksymalna prędkość gałki ocznej w fazie poszukiwania)
- *czas pozycjonowania* (czas pomiędzy zauważeniem nowego celu - spadek prędkości gałki ocznej, a skupieniem na nim uwagi - spadek wariancji do wartości typowej dla koncentracji uwagi),
- *opóźnienie reakcji* (czas pomiędzy prezentacją nowego celu a rozpoczęciem poszukiwania)
- *dokładność pozycjonowania* (wyrażona za pomocą wariancji pozycji podczas koncentracji uwagi).

Wyznaczanie *oka dominującego* zostało oparte na analizie wariancji sygnału okoruchowego podczas fiksacji wzroku. Proces wyznaczania niekiedy kończy się niepowodzeniem, ale są to sytuacje, kiedy rezultat wyznaczania nie ma znaczenia w aspekcie śledzenia punktu koncentracji uwagi. Być może z neurologicznego punktu widzenia, brak oka dominującego jest wariantem normy fizjologicznej. Ciekawym spostrzeżeniem jest natomiast częsta przewaga oka lewego, co z pewnością uzasadnia dalsze badania dotyczące percepcji dwuoczonej i podziału ról w procesie pozyskiwania informacji wizualnej.

2.4. Parametry ścieżki wzrokowej a strumień danych pobieranych ze sceny

Optymalna rozdzielczość percepcji wzrokowej jest ograniczona do niewielkiego obszaru siatkówki, a czas pozyskania całości informacji z rzutowanego obszaru sceny nie przekracza zwykle 500 ms, nawet jeżeli wzrok pozostaje dłużej skupiony nieruchomo. Z drugiej strony, obrazy prezentowane bardzo krótko (0.01 - 60 ms) są najpierw zapamiętywane w postaci biochemicznej na siatkówce oka, a następnie wstępnie przetwarzane i transmitowane przez nerw wzrokowy. Dzięki tej własności sekwencja obrazów prezentowana wystarczająco szybko jest postrzegana jako obraz ruchomy.

Powyższe zjawiska ograniczają zakres proporcjonalnej zależności czasu patrzenia i ilości pobranej informacji wizyjnej do przedziału 100-500 ms. Określenie tej zależności było przedmiotem eksperymentu wizualnego polegającego na pokazach przypadkowych sekwencji liter i cyfr (o długości 3-10 znaków) w ograniczonym czasie (100-800 ms). Zadaniem trzynastu obserwatorów (mężczyźni 18-35 lat) było wypowiedzenie zapamiętanej sekwencji natychmiast po pokazie. Ich wypowiedzi były rejestrowane, a poprawność oceniana na podstawie porównania z wzorcem. Następnie przeprowadzono analizę korelacji oceny poprawności z czasem pokazu oraz długością sekwencji.

3. Rezultaty

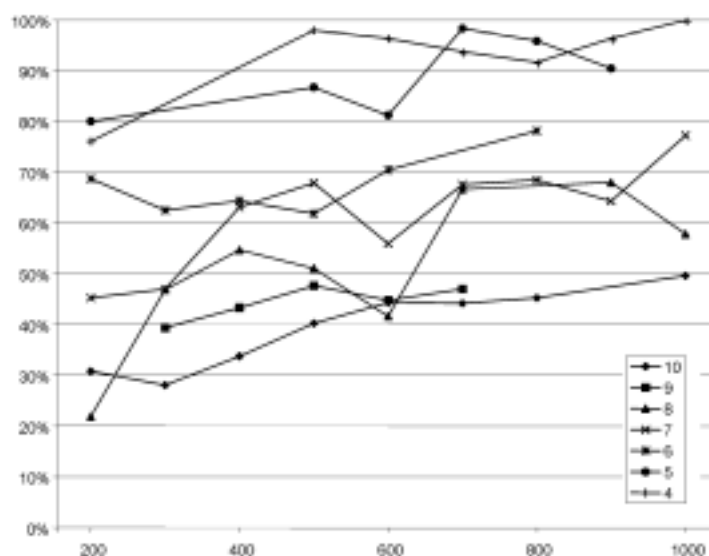
Podstawowymi rezultatami opisywanego eksperymentu było stwierdzenie korelacji pomiędzy poprawnością rozpoznania ciągu znaków a czasem prezentacji i długością ciągu. Pomimo znacznych różnic międzyosobniczych, w przypadku wszystkich obserwatorów wydłużenie czasu prezentacji powodowało poprawę wiarygodności odczytanej informacji. Wartości średnie przedstawia tabela 1. Tabela 2 przedstawia zależność poprawności odczytu ciągu od jego długości, natomiast rysunek 2 przedstawia zależność poprawności odczytu od czasu prezentacji dla ciągów znaków o różnej długości.

Tab. 1. Poprawność rozpoznania ciągu znaków w zależności od czasu prezentacji

| Czas [ms] | 200 | 300 | 400 | 500 | 600 | 700 | 800 |
|-----------|-----|-----|-----|-----|-----|-----|-----|
| % rozpoz. | 44% | 37% | 45% | 55% | 50% | 56% | 67% |

Tab. 2. Poprawność rozpoznania ciągu znaków w zależności od długości ciągu

| Długość | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------|-----|-----|-----|-----|-----|-----|-----|
| % rozpoz. | 91% | 88% | 68% | 61% | 54% | 45% | 42% |



Rys. 2. Zależność czasu obserwacji i ilości pozyskanej informacji - poprawność odczytu ciągów znaków o różnej długości od czasu ich pokazu

4. Dyskusja i wnioski

Analiza ścieżki wzrokowej w kontekście zawartości obserwowanej sceny pozwala na wyodrębnienie elementów przyciągających uwagę obserwatora w pierwszej kolejności, a także na ilościową ocenę zawartości informacyjnej poszczególnych regionów. Ocena taka powinna być przeprowadzona w polach siatki pokrywającej scenę. Stosowanie rozdzielczości siatki odpowiadającej rozdzielczości obrazu jest nieuzasadnione z dwóch powodów:

- z uwagi na rozmiar strefy dokładnego ogniskowania obrazu na powierzchni siatkówki oka, oś optyczna gałki ocznej nie musi dokładnie odpowiadać punktowi zainteresowania obserwatora, a jedynie wyznacza oś stożka koncentracji uwagi,
- pikselowa rozdzielczość siatki prowadzi do mylnego wniosku, że pobieranie tej samej ilości informacji (barwa piksela) podlega znacznym wahaniom.

Ilościowy opis lokalnej istotności obrazu może być zrealizowany poprzez średni czas skupienia uwagi na jednostkę powierzchni [s/m^2]. W przypadku obrazów przedstawiających wykresy, bardziej właściwą jednostką w każdym z wymiarów jest jednostka osi wykresu (np. czasu).

Określenie typowego sposobu postrzegania określonej klasy obrazów na drodze eksperymentalnej prowadzi do utworzenia modeli perceptualistycznych. Mogą one być wykorzystane w wielu zastosowaniach praktycznych: analizy zawartości graficznej witryn internetowych i optymalizacji rozmieszczenia symboli o priorytetowym znaczeniu, nauczaniu efektywnego postrzegania obrazów (np. medycznych) oraz tzw. technik szybkiego czytania.

5. Podziękowania

Niniejsza praca została sfinansowana ze środków AGH, nr umowy 10.10.120.39. Autorzy dziękują studentom, pp. Łukaszowi Chornikowskiemu i Łukaszowi Czajczykowi za pełną zaangażowania współpracę przy uruchamianiu i testowaniu oprogramowania użytego w eksperymentach.

Bibliografia

1. Laar P., Heskes T., Gielen S., Task-dependent learning of attention. *Neural Networks*, Vol. 10, No. 6, 1997.
2. Mikrut Z., Augustyniak P., Estimation of execution time for tasks of objects counting and localisation using the OBER2 device. *IFMBE Proc. vol. 2*, 2002. pp. 144-145
3. Ober J., Hajda J., Loska J., Jamicki M., "Application of eye movement measuring system OBER2 to medicine and technology". *Proc. of SPIE*, vol. 3061, part 1, 1997, pp. 327-333
4. Pelz J. B., Canosa R. "Oculomotor behavior and perceptual strategies in complex tasks", *Vision Research*, vol. 41, 2001, pp. 3587-3596
5. Yarbus A. L. "*Eye movements and vision*" Plenum Press, New York 1967