

# Proces uczenia i rozpoznawania wzorców na przykładzie systemu śledzenia ruchów gałki ocznej

**Krzysztof MURAWSKI, Witold ŻORSKI**

Instytut Teleinformatyki i Automatyki WAT,  
ul. Gen. S. Kaliskiego 2, 00-908 Warszawa  
k.murawski@ita.wat.edu.pl, w.zorski@ita.wat.edu.pl

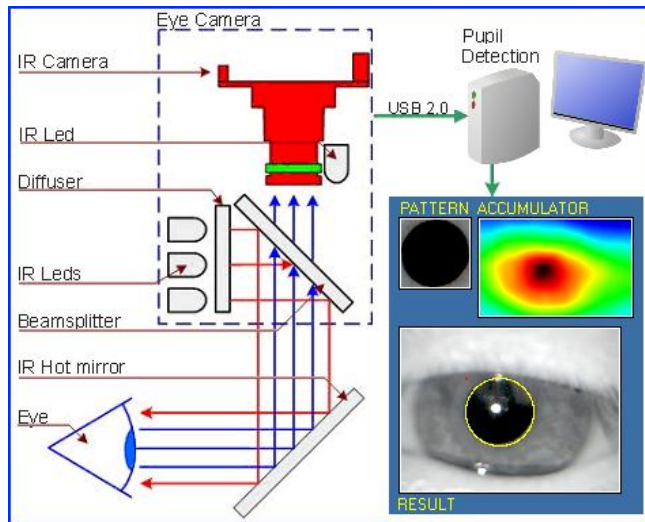
**STRESZCZENIE:** Przedstawiono metodę uczenia i rozpoznawania wzorców w systemie śledzenia ruchów gałki ocznej działającym w czasie rzeczywistym. Podstawowym zagadnieniem jest problem uczenia wzorców, który realizowany jest w oparciu o dynamicznie zarządzany zbiór wzorców. Zbiór wzorców jest adaptacyjny i w razie potrzeby nowy wzorec może zostać do niego dodany, aby następnie ewoluować w procesie dobierania wzorców do obiektów. Jako metodę dopasowywania wzorców do obiektów zastosowano transformatę Hougha dla wzorców nieregularnych. W celu zagwarantowania wydajności proces dopasowywania wzorców jest realizowany na wstępnie zawężonym fragmencie obrazu wejściowego.

**SŁOWA KLUCZOWE:** śledzenie gałki ocznej, dobieranie wzorców, transformata Hougha.

## 1. System śledzenia ruchów gałki ocznej

Rozważany jest system widzenia komputerowego zaprojektowany do śledzenia ruchów gałki ocznej [3], [9]. System działa w czasie rzeczywistym a jego schemat poglądowy został zaprezentowany na rys. 1.

Część systemu odpowiedzialna za pozyskiwanie obrazu gałki ocznej wykonana została w postaci urządzenia zakładanego na głowę (rys. 2). Pozwala ono na rejestrowanie ruchów źrenicy oka. Gałka oczna może być oświetlana przez diody podczerwone (ELIR15-21C/L10, ELIR17-21C, L850-03UP) w zakresie kątów oświetlania wynoszących (odpowiednio dla zastosowanej diody) 160, 120 i 30 stopni.



Rys. 1. Schemat systemu śledzenia ruchów gałki ocznej



Rys. 2. Urządzenie odpowiedzialne za pozyskiwanie obrazu oka

Światło IR emitowane przez diody pada na dyfuzor (ang. *diffuser*). Rozprasza on długości fal z zakresu od 800 nm do 1200 nm. Rozproszone światło doprowadzono do dzielnika optycznego (ang. *beam splitter*) dzielącego wiązkę w stosunku 50/50. Połowa światła jest bezpowrotnie tracona, pozostała część pada na zwierciadło IR (ang. *hot mirror*), które kieruje je w stronę oka. Światło powracające z oka jest ponownie odbijane od zwierciadła, przechodzi przez dzielnik wiązki i filtr światła widzialnego, aby ostatecznie trafić do

systemu optycznego kamery.

W urządzeniu zastosowano kamerę OptiTrack SLIM:V100, która pozwala na rejestrację obrazów z maksymalną rozdzielczością 640x480 pikseli oraz maksymalną częstotliwością 100 klatek na sekundę. Diody są umieszczone na płycie iluminatora w taki sposób, aby możliwe było otrzymanie zarówno obrazu „ciemnej” (ang. *dark pupil*) jak i „jasnej” (ang. *bright pupil*) źrenicy [11].

Podstawowym zagadnieniem w systemie jest problem rozpoznawania pozycji źrenicy w obrazie przechwyconym przez kamerę. Pozycja źrenicy w obrazie może być wyznaczana za pomocą różnych metod. Przykładowo, w [6] proponowane jest w tym celu śledzenie konturu (ang. *contour tracker*) w warunkach kontrolowanego środowiska, przy odpowiednich ustawieniach dla podczerwieni. W [7] zaprezentowano metodę trójwymiarowej videookulografii (ang. *videoculography*), pozwalającą na pomiar pionowych, poziomych i skrętnych ruchów oka wraz z pomiarem średnicy źrenicy.

Większość algorytmów, które można napotkać, stosuje techniki analizy obrazu. Dość popularne są metody oparte o różne warianty transformaty Hougha [4], [14], analizę tekstur [10], analizę różnicową obrazów [11], poszukiwanie miejsc o zmiennej jasności (ang. *corners*) [5], czy dopasowywanie prymitywów graficznych (okręgi, elipsy) [1], [13]. Rys. 3 ukazuje rezultaty dla przykładowych metod weryfikowanych w [12] (na zdjęciach widoczny jest charakterystyczny błysk w źrenicy – ang. *glint*).

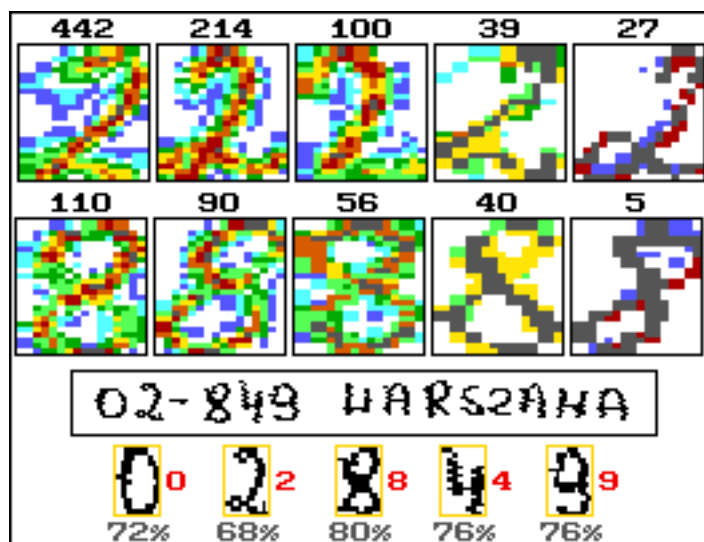


Rys. 3. a) metoda PCCM (ang. *pupil centre calculation method*) bez etykietowania, b) metoda PCCM z etykietowaniem, c) operator Daugmana

W przypadku metody Hougha najczęściej stosowana jest transformata Hougha dla okręgów (ang. *Circle Hough transform*). W niniejszym opracowaniu autorzy proponują jednak uogólnioną transformatę Hougha (ang. *Generalised Hough transform*) [2], która ma zastosowanie w przypadku obiektów nieregularnych, których nie można opisać prostym równaniem parametrycznym. Metoda jest dość złożona obliczeniowo, więc konieczne będzie wstępne ograniczenie obszaru obrazu wejściowego na drodze analizy histogramowej [15], [17]. Ze względu na różnorodność źrenicy powstała idea stworzenia zbioru wzorców oraz reguł do zarządzania nim.

## 2. Koncepcja zbioru wzorców nieregularnych

Koncepcja budowy zbioru wzorców nieregularnych ma swoją genezę we wcześniejszych doświadczeniach autora w zakresie rozpoznawania kodów pocztowych [15]. W przypadku rozpoznawania ręcznie pisanych kodów pocztowych pojawia się problem różnych charakterów pisma. Pomysł polegał na stworzeniu zbioru wzorców dla każdego znaku (cyfry) i użycia odpowiednich reguł zarządzania takim zbiorem. Jedną z podstawowych była zasada niewielkiej modyfikacji wzorca przez obiekt, który został tym wzorcem rozpoznany. Inna reguła polegała na promocji lub degradacji wzorców (włączając wykluczenie) jako konsekwencję ich użyteczności.



Rys. 4. Przykładowe zbiory wzorców dla cyfr ‘2’ i ‘8’ oraz rezultat rozpoznawania kodu

Rys. 4 ukazuje zbiory wzorców dla dwóch cyfr wraz z przykładowym wynikiem rozpoznania kodu pocztowego<sup>1</sup>. Liczby nad wzorcami to liczniki prawidłowego doboru danego wzorca. Jeśli dany wzorec został dopasowany do obiektu z wystarczająco dużą dokładnością (określoną procentowo na rys. 4), to taki wzorec podlegał modyfikacji na drodze delikatnego „odciśnięcia” przez obiekt, a jego licznik był zwiększany o 1, co zapewniało wzorcowi ewolucję i awans.

Przytoczony przykład jest intuicyjny i pokazowy, stanowi więc dobre wyjaśnienie ogólnej idei tworzenia zbioru wzorców żenic w przypadku systemu

<sup>1</sup> Rozpoznanie kodu pocztowego wymagało wcześniejszej lokalizacji linii z kodem pocztowym, podziału kodu na znaki oraz odpowiedniej normalizacji znaków.

śledzenia ruchów gałki ocznej. Niestety w przypadku źrenic zbiorów wzorców nie będzie tak atrakcyjny (rys. 5), chociaż reguły zarządzania nim będą podobne.

### 3. Zbiór wzorców i zarządzanie

W prezentowanym systemie identyfikacja obiektów jest prowadzona przy użyciu wcześniej uczonych wzorców. Obrazy wejściowe i wzorce są w poziomach szarości, więc na wstępie należy zdefiniować obraz w poziomach szarości, obiekt występujący w obrazie oraz wzorzec w poziomach szarości dla rozważanego systemu widzenia komputerowego.

#### Definicje

**Obraz w 256 poziomach szarości** jest zbiorem punktów (pikseli), które posiadają wartość (odcień) ze zbioru  $\{0, \dots, 255\}$ , opisanym następująco:

$$I_G : D \rightarrow \{0, \dots, 255\}, \quad \text{gdzie } D = [1, \dots, K] \times [1, \dots, L] \subset N^2. \quad (1)$$

**Obiektem** w obrazie  $I_G$  jest dowolny fragment tego obrazu, który można przedstawić w postaci:

$$Q_G : D_Q \rightarrow \{0, \dots, 255\}, \quad \text{gdzie } D_Q \subset D. \quad (2)$$

Uwaga: Utożsamienie obiektu z obrazem jest konsekwencją zbioru wartości przyjmowanych przez funkcję  $I_G$ .

**Wzorzec**  $M_p$  jest obrazem (macierzą kwadratową) o rozmiarze  $N_p \times N_p$ , którą można opisać następująco:

$$M_p : D_p \rightarrow \{0, \dots, 255\}, \quad \text{gdzie } D_p = [1, \dots, N_p] \times [1, \dots, N_p] \subset N^2. \quad (3)$$

Proces uczenia wzorców i zarządzania ich zbiorem może być dla omawianej metody przedstawiony w formie podstawowych reguł.

#### REGUŁY uczenia i zarządzania zbiorem wzorców

Reguła 1: System rozpoczyna pracę z jednym wzorcem ręcznie pozyskanym z obrazu wejściowego. Wzorcowi przypisana zostaje wartość licznika użycia równa 1.

Reguła 2: Jeśli wystąpi problem z rozpoznaniem źrenicy, wówczas następny wzorzec jest ręcznie wprowadzany do zbioru wzorców. Nowy wzorzec otrzymuje środkową pozycję w hierarchii wzorców, co zostaje wymuszone przez nadanie odpowiedniej wartości licznika użycia. Zbiór wzorców ma z góry ustaloną liczebność i jeśli nie ma już wolnego miejsca dla nowo wprowadzanego wzorca, to wzorzec z najmniejszym licznikiem jest usuwany ze zbioru.

Uwaga: „problem z rozpoznaniem źrenicy” występuje wówczas, gdy

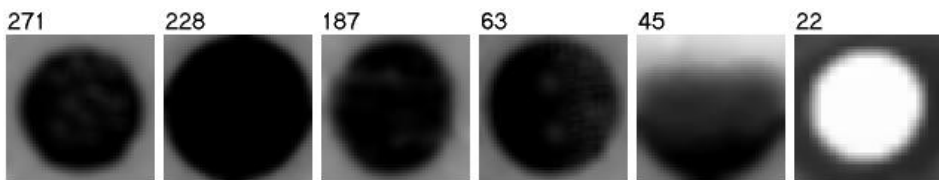
rezultat rozpoznania wskazuje na występowanie źrenicy w obrazie oka nieprzysłoniętej powieką, co można określić na podstawie wartości maksimum w akumulatorze (rys. 7).

**Reguła 3:** Jeśli wzorec  $M_P$  został dopasowany do obiektu  $Q_G$  w obrazie wejściowym (tj. wartość maksymalna w akumulatorze (9) przekroczyła odpowiednio dobrany próg), to wartość licznika użycia jest zwiększana o 1, a wzorec podlega następującej modyfikacji:

$$M_P^* = v \cdot M_P + (1 - v) \cdot Q_G, \quad (4)$$

gdzie wartość  $v \in (0,1)$ ; dobre rezultaty osiągnęte są w okolicy 0,95.

Przykładowy zbiór wzorców, wypracowany dla kilkunastu osób i setek obrazów video, jest przedstawiony na rys. 5. Liczby powiązane z wzorcami są wartościami wspomnianych liczników użycia. Zbiór zawiera odpowiedni wzorec dla przypadku częściowo przysłoniętej przez powiekę źrenicy oraz wzorec dla przypadku jasnej źrenicy (ang. bright pupil). Wzorce te nie są zbyt często używane, więc są dodatkowo chronione przed usunięciem ze zbioru wzorców (wyjątek do reguły 2).



Rys. 5. Przykład zbioru wzorców (jasna źrenica została uwzględniona)

#### 4. Uogólniona transformata Hougha

**Uogólniona transformata Hougha**<sup>2</sup>  $H(x_T, y_T, \alpha, s)$ , uwzględniająca translację o wektor  $[x_T, y_T]$ , obrót o kąt  $\alpha$  oraz skalowanie wyrażone przez  $s$ , dla obrazów w poziomach szarości  $I_G(x, y)$ , w procesie dopasowywania wzorca nieregularnego  $M_P$  może być zapisana następująco:

$$H(x_T, y_T, \alpha, s) = \sum_{(x_i, y_i) \in M_P} h(x_i, y_i, x_T, y_T, \alpha, s), \quad (5)$$

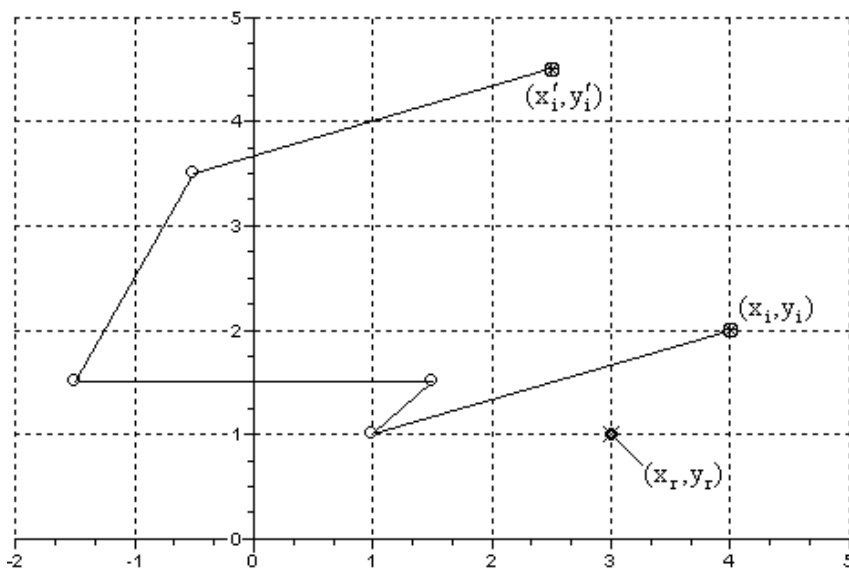
<sup>2</sup> Podana definicja odnosi się do ogólnego przypadku, który jest niezwykle złożony obliczeniowo i praktycznie nigdy nie stosowany.

gdzie<sup>3</sup>

$$h(x_i, y_i, x_T, y_T, \alpha, s) = 255 - |I_G(x'_i, y'_i) - M_P(x_i, y_i)|, \quad (6)$$

a wartości  $x'_i, y'_i$  są wyznaczone ze wzoru

$$\begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \begin{bmatrix} s \cos(\alpha) & -s \sin(\alpha) & -s x_r \cos(\alpha) + s y_r \sin(\alpha) + x_T + x_r \\ s \sin(\alpha) & s \cos(\alpha) & -s x_r \sin(\alpha) - s y_r \cos(\alpha) + y_T + y_r \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}. \quad (7)$$



**Rys. 6. Skalowanie, obrót i translacja ( $s = 1.5, \alpha = \pi / 2, x_T = 1, y_T = 2$ ) elementu  $(x_i, y_i)$  należącego do wzorca  $M_P$  względem  $(x_r, y_r)$**

Można zauważyć, że implementacja wzorów (5-7) wymaga określenia zakresu parametrów  $x_T, y_T, \alpha, s$ . W przypadku rozważanego systemu śledzenia ruchów gałki ocznej w eksperymentach brało udział kilkanaście osób i okazało się, że wystarczającym jest uwzględnienie tylko translacji  $x_T, y_T$ , co pozwala zdefiniować przestrzeń parametrów  $P = [1, \dots, K - N_p + 1] \times [1, \dots, L - N_p + 1]$ . Zysk ten wynika z charakteru obiektu rozpoznawanego, jakim jest źrenica oka.

<sup>3</sup> Funkcja podobieństwa (6) może być wprowadzona na różne sposoby, ale przedstawiony tutaj bezpośrednio wskazuje jak postępować w przypadku obrazów w poziomach szarości. Wartość 255 jest wprowadzona, aby rozważać maksimum a nie minimum w (9) i powinna być usunięta w implementacji programowej.

Wielkość źrenicy (problem skali) i występujące zniekształcenia zostały skompensowane przez zbiór wzorców. Niemniej jednak prace z bardziej zaawansowanym systemem 3D mogą wymagać uwzględnienia przynajmniej wąskiego zakresu dla kąta obrotu.

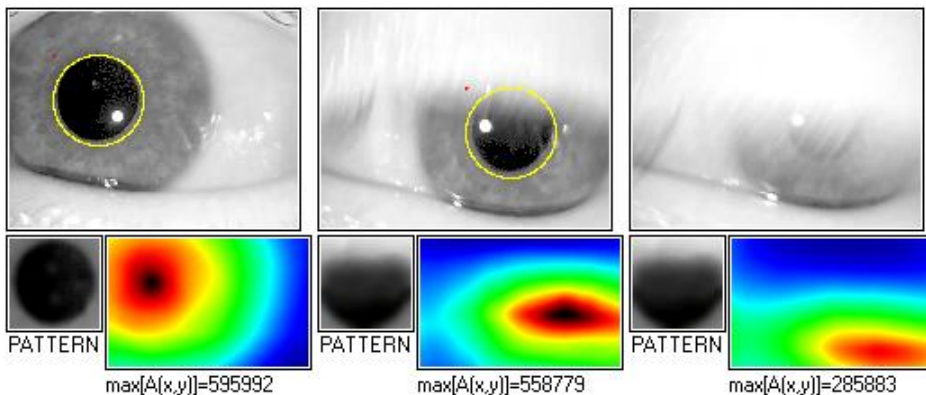
Koncepcja akumulatora jest ściśle związana z transformatą Hougha i wynika ze wspomnianej konieczności określenia przestrzeni parametrów. Akumulator  $A$  jest macierzą otrzymywaną jako rezultat transformacji obrazu  $I_G$ , co można zapisać następująco:

$$H : I_G \rightarrow A(x_T, y_T). \quad (8)$$

Mając obliczony akumulator dla obrazu  $I_G$  przy danym wzorcu  $M_P$  zasadniczym zagadnieniem jest wyznaczenie współrzędnych maksimum w akumulatorze

$$A(k_0, l_0) = \max_{(k,l) \in P} A(k, l), \quad (9)$$

co pozwala określić pozycję  $(k_0, l_0)$  obiektu (źrenicy) w obrazie.



Rys. 7. Detekcja źrenicy w przykładowych obrazach

Rys. 7 ukazuje przykładowe obrazy wejściowe, zastosowane wzorce i otrzymane akumulatory. Rezultat rozpoznawania (obiekt) jest zaznaczony okręgiem w obrazie. Czas obliczeń wynosił około 20 ms na ramkę, przy zastosowaniu procesora Core i5 2500 (3.3 GHz). Prezentowane rezultaty potwierdzają, że metoda jest obiecująca; pozwala nawet na prawidłowe wykrycie źrenicy częściowo przysłoniętej i odrzucenie (na podstawie zbyt małej wartości maksimum akumulatora) obrazu zamkniętego oka.



## 5. Rozpoznawanie orientacji obiektu

W celu zwiększenia wydajności omawianej metody można wykorzystać analizę histogramową. Histogram jest zasadniczą charakterystyką obrazów w poziomach szarości, posiada też bardzo ważną własność jaką jest niezmienność w przypadku obrotu obiektu. Histogram obrazu w poziomach szarości będzie rozumiany jako funkcja przypisująca danemu poziomowi szarości (od 0 do 255) liczbę pikseli o tej wartości w obrazie

$$\Phi : \{0, \dots, 255\} \rightarrow N. \quad (10)$$

W rozpatrywanym zagadnieniu histogram danego wzorca  $M_P$  wyznaczany jest jeden raz a następnie porównywany z histogramem każdego fragmentu obrazu  $I_G$ .

**ALGORYTM** – analizy histogramowej

**Krok 1:** Wyznaczenie histogramu  $\Phi_P$  dla danego wzorca  $M_P$ .

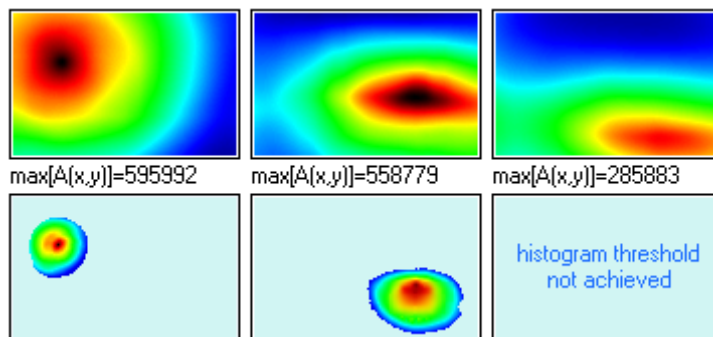
**Krok 2:** Wyznaczenie histogramów  $\Phi_{Q(i,j)}$  fragmentów  $Q_G^{I(i,j)}$  o rozmiarze  $N_P \times N_P$  w obrazie  $I_G$ , gdzie:  $i = 1, \dots, K - N_P + 1$ ,  $j = 1, \dots, L - N_P + 1$ .

**Krok 3:** Porównanie każdego histogramu  $\Phi_{Q(i,j)}$  z histogramem  $\Phi_P$ :

$$d_{(i,j)} = \frac{100\%}{2 \cdot N_P^2} \sum_{k=0}^{255} |\Phi_{Q(i,j)}(k) - \Phi_P(k)|. \quad (11)$$

Uwaga: W przypadku porównywania identycznych histogramów uzyskamy 0% a w najgorszym możliwym przypadku 100%.

**Krok 4:** Jeśli wartość  $d_{(i,j)}$  jest większa niż przyjęta wartość progowa  $d_{threshold}$ , to  $(i, j)$  jest wykluczane przy obliczaniu akumulatora.



Rys. 8. Zysk z analizy histogramowej (por. rys. 7)

Zastosowanie analizy histogramowej redukuje czas obliczeń transformaty Hougha zazwyczaj o ponad 50%. Należy podkreślić, że czas obliczeń związanych z analizą histogramową można zaniedbać w porównaniu do czasu obliczania transformaty Hougha. Rys. 8 ukazuje zysk wypracowany na etapie analizy histogramowej (obrazy i wzorce są takie same jak w przypadku z rys. 7).

## 6. Podsumowanie

Podczas eksperymentów okazało się, że metoda jest dość szybka (20 ms na ramkę) jak również skuteczna (>90% prawidłowo rozpoznanych obrazów). Obrazy przedstawiające oko przysłonięte powieką były łatwo odrzucane ze względu na małe wartości maksymalne w otrzymanywanych akumulatorach (rys. 7). Okazało się również, że transformata Hougha jest odporna na tzw. „błysk w źrenicy” (ang. *glint*).

Należy wyjąć, że proces zarządzania wzorcami nie jest aż tak prosty, jak wynikałoby to z przedstawionych reguł. W trakcie eksperymentów konieczne było wprowadzenie dodatkowych ograniczeń (nadal rozwijanych), które przykładowo pozwalały uniknąć niestabilności procesu uczenia wzorców.

Czas obliczeń został efektywnie zredukowany na drodze wprowadzenia analizy histogramowej. Dalsza redukcja jest możliwa poprzez zastosowanie nierównomiernego podziału w przestrzeni parametrów (tzw. szybka transformata Hougha). Rozważane jest zastosowanie przedstawionej metody w przypadku systemu śledzenia 3D, gdzie odpowiednie przyspieszenie obliczeń będzie osiągnięte na drodze wykorzystania technologii CUDA [16], dostępnej nawet w laptopach. Łatwo zauważyć, że w zaprezentowanej metodzie liczba niezależnych wywołań transformaty Hougha jest proporcjonalna do liczby pikseli w obrazie wejściowym, co pozwala na masywne zrównoleglenie obliczeń z wykorzystaniem procesorów graficznych [8].

## Literatura

- [1] AJDARI R. A., FAEZ K., QARAGOZLOU N., et. al., *Fast Circle Detection Using Gradient Pair Vectors*. Proceedings of the Seventh International Conference on Digital Image Computing: Techniques and Applications, CSIRO Publishing, 2003, pp. 879-888.
- [2] BALLARD D. H., *Generalizing the Hough Transform to Detect Arbitrary Shapes*. Readings in Computer Vision: Issues, Problems, Principles, and Paradigms. Los Altos, CA. 1987, pp. 714-725.

- [3] FENG J., GUIMING H., *Real-Time Eye Detection and Tracking under Various Light Conditions*, Data Science Journal, Vol. 6, 2007.
- [4] GAO J., et al., *Application of Hough transform in eye tracking and targeting*. 9th International IEEE Conference on Electronic Measurement & Instruments, ICEMI '09, Beijing, China, 2009, pp. 751-754.
- [5] GUPTA P., MEHROTRA H., RATTANI A., CHATTERJEE A., KAUSHIK A. K., *Iris Recognition using Corner Detection*. 23rd International Biometric Conference, Montreal, Canada, 2006
- [6] HANSEN D., PECE A., *Eye tracking in the wild*. Computer Vision and Image Understanding, vol. 98, no. 1, 2005, pp. 155-181.
- [7] KIM S. C., NAM K. C., LEE W. S., KIM D. W., *A new method for accurate and fast measurement of 3D eye movements*. Medical Engineering & Physics, Vol.28, pp. 82-89.
- [8] KIRK B., HWU W., *Programming Massively Parallel Processors: A Hands-on Approach*. NVIDIA Corporation, Morgan Kaufmann Publishers 2010.
- [9] LI D., PARKHURST D. J., *Starburst: A robust algorithm for video-based eye tracking*, Image Rochester NY, 2005.
- [10] MA L., TAN T., WANG Y., ZHANG D., *Personal identification based on iris texture analysis*. IEEE Transactions Pattern Analysis and Machine Intelligence, Vol. 25, 2003, pp. 1519-1533.
- [11] MORIMOTO C. H., KOONS D., AMIR A, FLICKNER M., *Frame-rate pupil detector and gaze tracker*. Proceedings of the ICCV 99 Frame-rate workshop, 1999.
- [12] MURAWSKI K., *Method for determining the position of the pupil for eyetracking applications*. Proceedings of the 15th IEEE International Conference MMAR 2010, pp. 356-362.
- [13] SHAPIRO L. G., STOCKMAN G. C., *Computer Vision*. Prentice Hall, 2001.
- [14] TOENNIES K. D., BEHRENS F., AURNHAMMER M., *Feasibility of Hough-Transform-based Iris Localisation for Real-Time-Application*. Proceedings 16th International Conference on Pattern Recognition (ICPR), 2002 pp. 1053-1056.
- [15] ZORSKI W., FOXON B., BLACKLEDGE J., Turner M., *Irregular Pattern Recognition Using the Hough transform*. Machine Graphics & Vision, 9, 2000, pp. 609-632.
- [16] ZORSKI W., MAKOWSKI M., *Zastosowanie technologii CUDA w rozpoznawaniu wzorców nieregularnych*. Biuletyn WAT, Vol.LX, 4/2011, str. 265-279.
- [17] ZORSKI W., *The Hough Transform Application Including Its Hardware Implementation*. Advanced Concepts for Intelligent Vision Systems: Proceedings of the 7th International Conference, Lecture Notes in Computer Science, Springer-Verlag Vol. 3708/2005, pp. 460-467.

## **Process of patterns learning and matching on example of an eye tracking system**

**ABSTRACT:** This paper presents a method of patterns learning and matching in a real-time computer vision system designed for eye tracking. The main issue is the problem of patterns learning that requires creation of a set of patterns with some kind of dynamic management. The set of patterns is adaptive in that way that a new pattern can be added to the set and next undergoes an evolution during the process of matching. The Hough transform for irregular patterns play the role of the matching method. In order to guarantee the performance the process of patterns matching is carried out on a preliminary selected narrow fragment of the input image.

**KEYWORDS:** eye tracking, pattern matching, Hough transform.

*Praca wpłynęła do redakcji: 8.10.2012*