AUTOMATYKA • 2006 • Tom 10 • Zeszyt 3

Piotr Pawlik*, Daniel Iwaniec**, Michał Iwaniec**

Analiza obrazu z kamery jako podstawa interfejsu człowiek niepełnosprawny-komputer***

1. Wprowadzenie

Problem umożliwienia lub chociaż usprawnienia osobom niepełnosprawnym samodzielnej pracy z komputerem został dostrzeżony już w początkowych etapach prac nad graficznymi interfejsami użytkownika systemów operacyjnych. Przykładem mogą być „Ułatwienia dostępu", które pojawiły się np. w systemie Microsoft Windows 95. Oczywiście oprócz ułatwień dostarczanych z systemem powstawało coraz więcej dedykowanych aplikacji wspomagających osoby niepełnosprawne w pracy z komputerem. Należy tu jednak rozróżnić aplikacje wspomagające pracę od aplikacji umożliwiających pracę na komputerze. Wspomniane „Ułatwienia dostępu", zgodnie zresztą ze swoją nazwą, mogą tylko wspomóc osoby lekko niepełnosprawne. Umożliwienie pracy na komputerze osobom z bezwładnością (lub po amputacji) kończyn wymagało stworzenia aplikacji praktycznie zrywających z tak podstawowymi urządzeniami wejścia/wyjścia, jak klawiatura i mysz (nie wspominając o urządzeniach bardziej specjalizowanych). W takich przypadkach jedynymi możliwymi (powszechnie dostępnymi) sensorami komputera stają się kamera i mikrofon. Przy czym istotne wydaje się wyróżnienie zakresu stosowalności obu tych kanałów komunikacji. Sterowanie komputera za pomocą głosu skupia się na wydawaniu poleceń i rozpoznawaniu mowy w celach tworzenia dokumentów. Natomiast droga wizyjna jest wykorzystywana do zastąpienia urządzeń wskazujących. Niniejszy artykuł skupia się na tym drugim aspekcie komunikacji człowieka niepełnosprawnego z komputerem.

2. Istniejące aplikacje wizyjnej komunikacji z komputerem

Zastąpienie urządzenia wskazującego kamerą oznacza pozyskanie obrazu osoby sterującej kursorem i analizą mającą na celu przekształcenie informacji zawartej w tym obrazie

* Katedra Automatyki, Akademia Górniczo-Hutnicza w Krakowie; piotrus@agh.edu.pl
** Instytut Politechniczny, PWSZ w Tarnowie
*** Praca powstała w ramach badań własnych – umowa AGH nr 10.10.120.39

na komendy przemieszczające kursor. Zadanie to może być realizowane na wiele sposobów, jednakże sprowadzają się one do dwóch nurtów:

1) wyodrębnienia z obrazu punktu (lub obszaru), którego ruch jest zamieniany na ruch kursora;
2) sterowania kursorem na podstawie ruchu gałki ocznej.

Poniżej omówiono przykładowe aplikacje dla obu typów rozwiązań.

2.1. CameraMouse

Jest to projekt rozwijany przez Boston College. Jego celem jest pomoc osobom, które nie mają możliwości poruszania żadną częścią ciała w sposób umożliwiający używanie tradycyjnego wskaźnika, jakim jest mysz. System wykorzystujący prostą kamerę analizuje ruchy głowy (lub jej części, np. nos, podbródek, usta) i odpowiednio przesuwa kursor na ekranie monitora. Kamera powinna być zogniskowana na twarz, gdyż znajdowany jest nos, ewentualnie podbródek osoby sterującej. Zmiany położenia tych części twarzy są podstawą obliczenia pozycji kursora na ekranie. Zdarzenia kliknięcia przyciskiem myszy są wywoływane, gdy użytkownik zatrzyma kursor na krótki interwał czasowy w jednym miejscu. Stanowi to pewną wadę rozwiązania, gdyż zatrzymanie kursora na dłuższy czas nie musi oznaczać chęci naciśnięcia klawisza myszy [1, 5].

2.2. Quick Glance

Quick Glance umożliwia osobom niepełnosprawnym obsługę komputera, wykorzystując ruch gałki ocznej obserwowanej przez kamerę. Zdarzenie kliknięcia myszy można zrealizować przez krótkie mrugnięcie oka lub zatrzymanie wzroku na pewien, krótki czas w jednym miejscu. Rozwiązanie to może być stosowane w pomieszczeniach o różnorodnej jakości oświetlenia. Z faktu, iż kamera w zupełności skupia się na oku, wynika, że nie jest dopuszczalne znaczne poruszanie głową. Zaletą takiego rozwiązania jest szybkość poruszania kursora. Do wad należy zaliczyć wspomaganie diodami podczerwonymi oraz fakt, iż wykorzystanie oka do pozycjonowania kursora wiąże się z koniecznością neutralizowania drgań, jakie wprowadza gałka oczna [4, 6].

3. Proponowana metoda oparta o rozpoznawanie wzorców

3.1. Analiza ograniczeń

W dotychczasowych opisach świadomie pominięto rozwiązania interfejsu człowiek-komputer bazujące na dodatkowych urządzeniach poza kamerą (lub zastępujące kamerę), zarówno ze względu na konieczność dodatkowego przygotowywania się do obsługi komputera (np. czujniki elektrookulograficzne), jak i z powodu wysokiego kosztu takich rozwiązań.

Rozwiązania wykorzystujące śledzenie twarzy zakładają ruchomość głowy. Przy rozległym paraliżu kończyn często poruszanie głową także jest ograniczone, co w tym wypadku dyskwalifikuje te metody. Możliwe jest śledzenie wybranego ruchomego punktu twarzy (np. podbródka), ale najczęściej jest to związane z kłopotliwym umieszczaniem na skórze dodatkowego markera. Z kolei obserwacja ruchu gałki ocznej jest wrażliwa na gwałtowne jej ruchy, oraz często bywa wspomagana diodami podczerwonymi, stanowiącymi dodatkowe wyposażenie.

Analiza powyższych ograniczeń zaowocowała propozycją rozwiązania niezależnego od ruchu gałki ocznej i nie skupiającego się na śledzeniu twarzy lub innych elementów obrazu. Istotne też było opracowanie rozwiązania nie wymagającego dodatkowego sprzętu lub wysokiej jakości (drogich) kamer. W proponowanym rozwiązaniu ruch kursora oraz kliknięcia myszy są odpowiedziami na „rozkazy mimiczne". Założono, iż system ma rozpoznawać proste komendy ruchu kursora („góra", „dół", „lewo", „prawo") oraz naciśnięcia przycisków myszy. Rozpoznawanymi „komendami" byłyby obrazy zróżnicowanych min użytkownika komputera. Ponieważ proponowany system miał w założeniach pracować na tanim sprzęcie w różnych warunkach oświetleniowych i być niezależny od ewentualnego poruszania się użytkownika, zdecydowano się użyć metody SIFT (*Scale Invariant Features Transform*), która wyróżnia się spełnieniem wszystkich powyższych wymagań oraz dużą niezawodnością.

3.2. Propozycja rozwiązania opartego o *Scale Invariant Features Transform*

Metoda SIFT zaproponowana przez Dawida G. Lowe z University of British Columbia [2, 3] jest algorytmem detekcji i opisu charakterystycznych punktów obrazu. Punkty uzyskane za jego pomocą charakteryzują się odpornością na obrót, zmianę skali, a także pewną odpornością na zmianę oświetlenia.

Metoda SIFT bazuje na stworzeniu piramidy obrazów o coraz mniejszej rozdzielczości. Obrazy tworzące piramidę są obrazami różnicowymi. Są one uzyskiwane w wyniku odjęcia dwóch obrazów powstałych przez przefiltrowanie obrazu początkowego filtrami Gaussa o różnych parametrach σ.

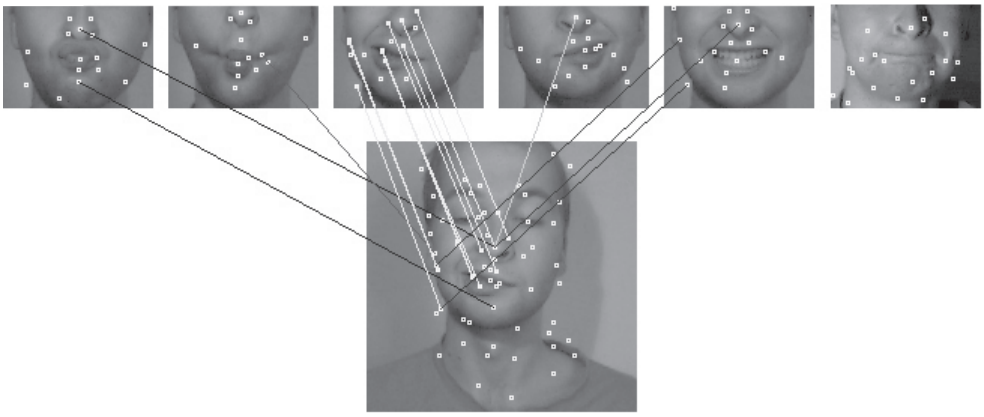
Na wstępie działania algorytmu na obrazie wyszukiwane są punkty charakterystyczne. Znajdowane są one jako ekstrema w obrazach tworzących piramidę. Następnie w otoczeniach ekstremów wyznacza się wartość i kierunek gradientu dla każdego punktu i tworzy się histogram orientacji tych gradientów względem gradientu w punkcie charakterystycznym. Tak wyznaczony histogram staje się deskryptorem cech. Porównując deskryptory na dwóch obrazach można znaleźć punkty wspólne tych obrazów.

Posiadając obrazy wzorcowe min można stwierdzić, czy występują one na obrazie z kamery – porównując deskryptory zawarte w obrazie z deskryptorami wzorców, wyszukuje się wzorzec o największej liczbie wspólnych punktów charakterystycznych. Jeżeli liczba punktów wspólnych nie przekracza zadanego progu, minę uznaje się za neutralną (brak dopasowania do jakiegokolwiek wzorca).

4. Weryfikacja metody

4.1. Aplikacja testowa

W celu weryfikacji przedstawionej metody, napisana została aplikacja do rozpoznawania i przetwarzania mimiki twarzy na ruchy kursora. Aplikacja pobiera obrazy z prostej kamery internetowej (typu QuickCam), który następnie poddaje się procesowi analizy za pomocą algorytmu SIFT. Wyznaczone na obrazie punkty charakterystyczne porównuje się z (uprzednio wprowadzonymi na etapie konfiguracji aplikacji) obrazami bazowymi przedstawiającymi miny (por. rys. 1) przypisane do odpowiednich ruchów kursora oraz do zdarzeń kliknięcia lewym i prawym klawiszem myszy. Po stwierdzeniu wystarczającego podobieństwa do jednego z obrazów bazowych następuje przesunięcie kursora lub wywołanie zdarzenia kliknięcia.



Rys. 1. Przykładowy wynik porównania obrazu z kamery (dół) z sześcioma wzorcami (góra)

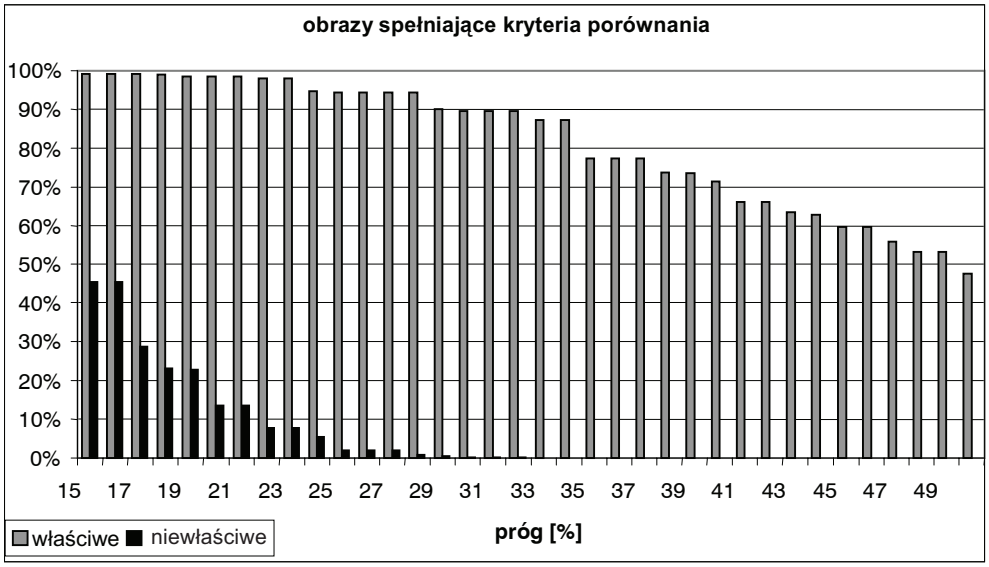
4.2. Test metody

Aby sprawdzić poprawność metody, wykonano testy na 360 obrazach zawierających miny. Każdy kolejny obraz wczytywany z kamery był porównywany do sześciu obrazów z minami bazowymi (rys. 1).

Aby obraz został uznany za podobny do obrazu bazowego, liczba porównań musiała przekroczyć zadany próg podobieństwa (tzn. liczba punktów charakterystycznych wspólnych dla obu obrazów musiała być większa od zadanej wartości progowej). Wartość tego progu wyznaczono doświadczalnie, co zostało zobrazowane na rysunku 2. Jest to procentowy wykres poprawnych i niepoprawnych rozpoznań w zależności od przyjętego progu podobieństwa. Jasny wykres słupkowy przedstawia procent obrazów o minie zgodnej z miną

bazową (które zostały prawidłowo zakwalifikowane), a ciemny pokazuje, ile obrazów zostało zakwalifikowanych niewłaściwe.

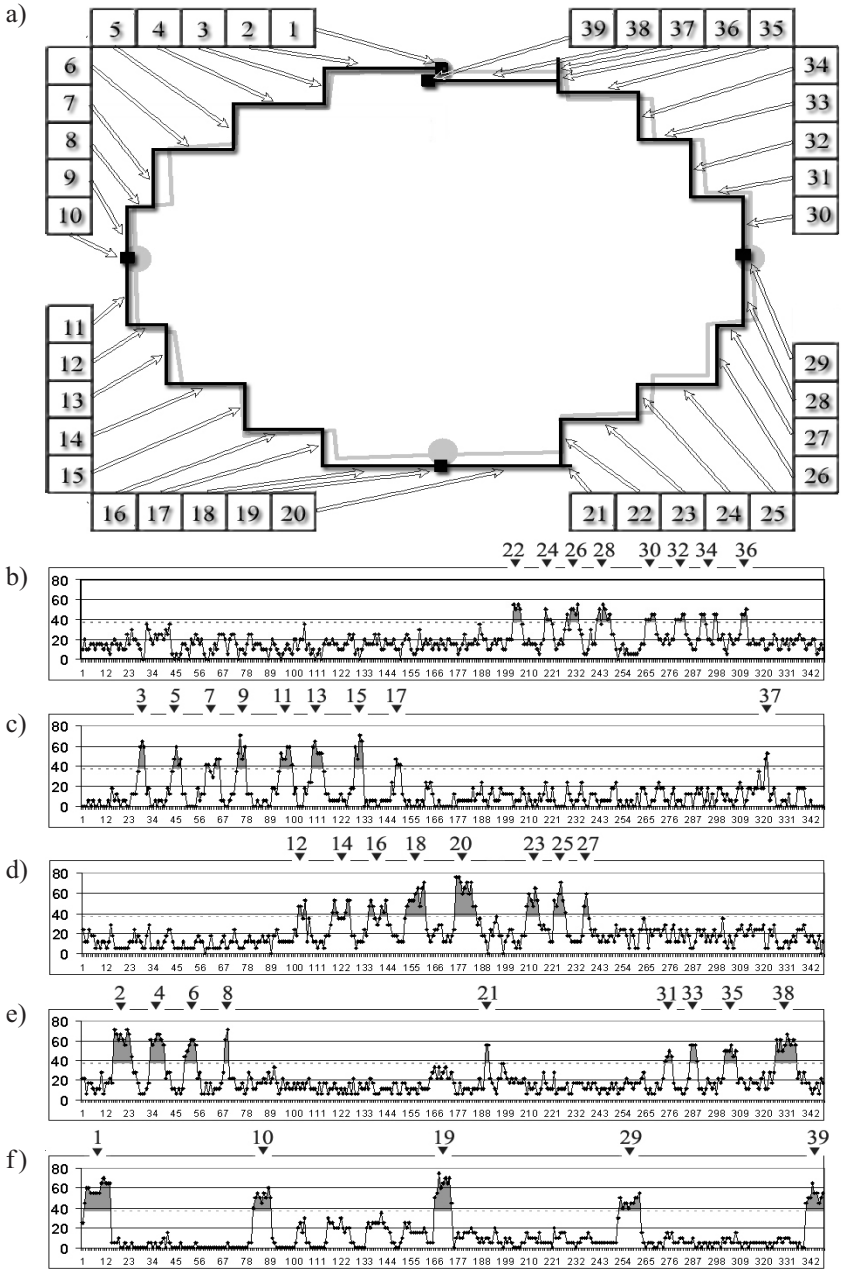
Analiza wyników sugeruje użycie progu podobieństwa w okolicach 35%. Jak widać, dla progu 33% nie wystąpiło żadne błędne dopasowanie, a 87% obrazów z minami zgodnymi z obrazem bazowym zostało poprawnie rozpoznane.



Rys. 2. Procentowy wykres odpowiedzi systemu w zależności od progu podobieństwa

4.3. Test aplikacji

W celu sprawdzenia działania aplikacji wykonano proste testy mające na celu sprawdzenie, jak skutecznie można poruszać kursorem i wywoływać kliknięcia przycisków myszy. Na ekranie narysowano szarą linią figurę o kształcie przypominającym elipsę (por. rys. 3), a następnie nawigowano kursorem wzdłuż jej krawędzi, wykonując „kliknięcia" na obszarach oznaczonych kółkami. Ruch kursora obrazują czarne linie. Strzałkami zaznaczono miejsca, w których znajdował się kursor w chwili pozyskania pierwszego i ostatniego obrazu testowego. Liczby wokół „elipsy" służą powiązaniu odcinków pokonywanych przez kursor z umieszczonymi poniżej wykresami. Na wykresach przedstawiono procentową liczbę zgodnych punktów charakterystycznych obrazu z punktami każdego z pięciu obrazów bazowych dla kolejnych obrazów z kamery. Przerywaną linią na każdym wykresie zaznaczono wartość przyjętego progu akceptacji (37%), a obszary zaciemnione powyżej tego progu akceptacji oznaczają rozpoznanie wzorca. Liczby nad wykresami odpowiadają liczbom reprezentującym odcinki pokonywane przez kursor. Wskazują one miejsca na wykresach, gdzie następowało rozpoznanie danej komendy.



Rys. 3. Śledzenie „elipsopodobnego” wzorca (a). Wykresy rozpoznań pięciu wzorców dla kolejnych 349 obrazów z kamery: b) góra; c) dół; d) prawo; e) lewo; f) kliknięcie lewym przyciskiem myszy
Objaśnienia w tekście

5. Podsumowanie

Wyniki testów, zarówno samej metody, jak i aplikacji na niej opartej potwierdzają, iż możliwe jest zastosowanie wzorców mimicznych jako ekwiwalentów komend wydawanych urządzeniu wskazującemu komputera. Co więcej, proponowana metoda posiada zalety umożliwiające nieuciążliwe stosowanie jej przez użytkowników o znacznym stopniu niepełnosprawności ruchowej bez ponoszenia wysokich kosztów zakupu specjalizowanego sprzętu. Dalsze prace powinny skupić się na dalszym przyspieszeniu działania algorytmu poprzez próbę optymalizacji metody SIFT pod kątem tej konkretnej aplikacji.

Literatura

[1] Betke M., Gips J., Fleming P.: *The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access For People with Severe Disabilities*. IEEE Transaction on Rehabilitation Engineering, 10, 1, 2002, 1–10

[2] Lowe D.G.: *Object recognition from local scale-invariant features*. International Conference on Computer Vision, 1999, 1150–1157

[3] Lowe D.G.: *Distinctive image features from scale-invariant keypoints*. International Journal of Computer Vision, 60, 2, 2004, 91–110

[4] Rasmusson D., Chappell R., Trego M.: *Quick Glance: Eye-Tracking Access To The Windows95 Operating Environment*. Technology And Persons With Disabilities Conference, 1999

[5] http://www.cameramouse.com

[6] http://www.eyetechds.com