

Paweł Gardziński,
Łukasz Kamiński,
Krzysztof Kowalak,
Sławomir Maćkowiak
Katedra Telekomunikacji Multimedialnej i Mikroelektroniki
ul. Polanka 3, 61-131 Poznań,
tel: + (48) 6653900, fax: + (48) 6653899,
email: {pgardzinski, lkaminski, kkwolak, smack}@multimedia.edu.pl
web: www.multimedia.edu.pl



Warszawa, 11-13 czerwca 2014

ROZPOZNAWANIE ZACHOWAŃ OSÓB Z WYKORZYSTANIEM PUNKTÓW MINIMUM I MAKSIMUM KRZYWIZNY OTOCZKI CIAŁA

Streszczenie: W artykule przedstawiono metodę automatycznego rozpoznawania wyuczonych zachowań osób z wykorzystaniem modelu opartego o graf ukierunkowany, podklasę dynamicznych sieci Bayesa. W proponowanym rozwiązaniu zastosowano jako punkty charakterystyczne punkty minimum i maksimum krzywizny otoczki ciała wyznaczonej sylwetki. Wyniki eksperymentalne potwierdzają wysoki stopień rozpoznawania zachowania.

1. WSTĘP

W ostatnich latach znacząco wzrosła liczba kamer rejestrujących zachowania osób w różnych dziedzinach naszego życia. Doprowadziło to do znacznego przyrostu ilości analizowanego materiału wizyjnego i w konsekwencji do zapotrzebowania na aplikacje służące między innymi opisowi zawartości treści. Znaczenie metod analizy zachowań osób stało się bardzo ważne wraz z rozwojem możliwości wykorzystania obrazu wizyjnego w systemach ochrony i dozoru wizyjnego. Szczególnie ważną grupą rozwiązań są zautomatyzowane systemy obserwacji obszarów ruchu pieszego i wykrywania niebezpiecznych zachowań. Tego rodzaju zadanie obserwacji nie jest jednak dobrze dopasowane do operatora systemu jakim jest człowiek, gdyż wymaga dokładnego skupienia przez długi okres czasu. Dlatego też, jest to motywacja do rozwijania zautomatyzowanych, inteligentnych systemów monitoringu opartego na obrazie ruchomym, które mogłyby pomóc użytkownikowi w procesie wykrywania i analizy ryzyka.

Modelowanie zachowania ludzi z wykorzystaniem grafów jest stosowane z dobrymi efektami do modelowania przede wszystkim złożonych zachowań osób i opisu zachowań osób w zatłoczonych miejscach publicznych. Modelowanie aktywności osób na podstawie sekwencji wizyjnych rejestrowanych z wykorzystaniem jednej kamery może być realizowane przez wiele różnych technik m.in. modele oparte o grafy prawdopodobieństwa (np. *sieci Bayesa* [1,2], *dynamiczne sieci Bayesa* [3,4], *sieci propagacji* [5]), modele semantyczne [6], sieci Petriego, podejścia syntaktyczne lub regulowe. Przegląd rozwiązań dotyczących sieci Petriego, sieci neuronowych i podejść syntaktycznych można znaleźć w opracowaniu [7].

Sieć Bayesa lub algorytm propagacji wierzeń jest acyklicznym kierunkowym grafem prawdopodobieństwa. Model składa się z węzłów, które reprezentują zmienne losowe (np. kolejne stany zdarzenia) i przejścia,

które reprezentują warunkowe zależności pomiędzy węzłami. Siła zależności jest parametryzowana przez prawdopodobieństwa warunkowe, które są przypisane do każdej grupy węzłów rodzic-dziecko w sieci. Taka sieć ma potężne możliwości w reprezentacji obserwacji związków przyczynowych, w związku z czym jest bardzo popularnym narzędziem do modelowania aktywności osób. Dla przykładu, sieci propagacji, podgrupa dynamicznych sieci Bayesa z możliwością modelowania wprost trwania interwałów czasowych została wykorzystana w [8] do analizy trwania podprzedziałów czasowych kilku równoległych strumieni zdarzeń.

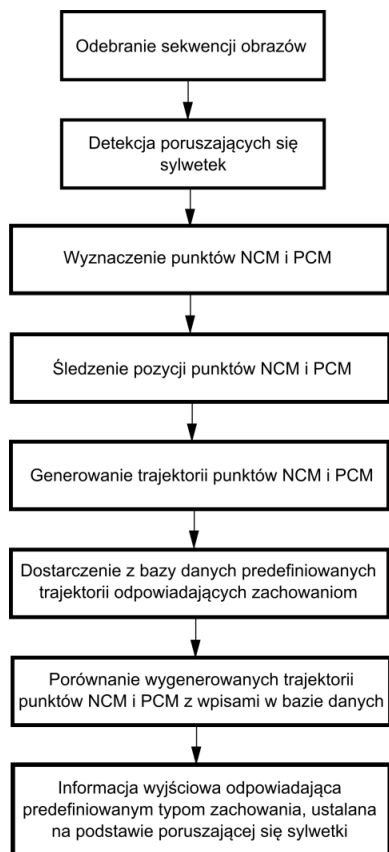
Artykuł podzielony jest na 4 główne punkty. Punkt 2 przedstawia cały system rozpoznawania zachowań osób wraz z wyjaśnieniem podstawowych bloków przetwarzania sygnału wizyjnego. Punkt 3 dostarcza szczegółowych wyjaśnień na temat punktów charakterystycznych NCM i PCM, których wykorzystanie w modelowaniu jest głównym tematem pracy. Punkt 4 przedstawia założenia eksperymentów i osiągnięte rezultaty dla przykładowego zachowania - wołania o pomoc. W punkcie 5 przedstawiono podsumowanie oraz sugestie dalszych badań w zakresie tematycznym pracy.

2. SYSTEM ROZPOZNAWANIA ZACHOWAŃ

Proponowane rozwiązanie pozwala na wykorzystanie w procesie detekcji punktów charakterystycznych leżących na konturze sylwetki człowieka. Zachowanie może być opisane jako zbiór trajektorii przemieszczania się punktów charakterystycznych podczas wykonywania danej czynności. Zbiór punktów charakterystycznych w danej chwili czasu definiuje pozę. Zbiór póz zdefiniowanych dla kolejnych punktów czasu tworzy deskryptor. Zbiór punktów definiujący pozę może mieć różną konfigurację dla różnych typów wykrywanego zachowania. Dla kolejnych ramek obrazu, pozycje punktów należących do zbioru są śledzone i tworzą trajektorię punktów.

Rysunek 1 przedstawia ogólny schemat proponowanego systemu rozpoznawania zachowań ludzi. Przedstawiony system wykorzystuje pojedynczą, stacjonarną kamerę.

Główną ideą prezentowanego rozwiązania jest śledzenie pozycji punktów minimum krzywizny (ang. Negative Curvature Minima, NCM) i punktów maksimum krzywizny (ang. Positive Curvature Maxima, PCM) (definicje podane w następnym punkcie),

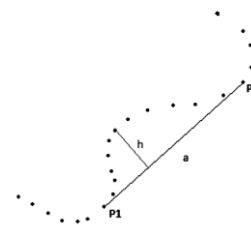


Rys. 1. Schemat ogólny systemu rozpoznawania zachowań.

na wykrytej ruchomej sylwetce człowieka. Pozycja punktów w zbiorze punktów w kolejnych ramach obrazu jest śledzona w celu wygenerowania trajektorii. Następnie trajektorie są porównywane z predefiniowanymi trajektoriami odpowiadającymi danemu zachowaniu. Porównanie jest przeprowadzone poprzez obliczenie odległości euklidesowej dla par odpowiadających sobie punktów. Każda trajektoria musi mieścić się w określonym z góry przedziale.

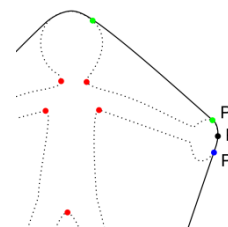
3. PUNKTY MINIMUM I MAKSYMUM KRZYWIZNY

Istnieje wiele różnych sposobów opisu kształtów w obrazach. Zaproponowane rozwiązanie wykorzystuje kontury obiektów wykrytych w scenie. Kontur można dobrze scharakteryzować wykorzystując tzw. minimum wklęsłości lub punkt minimum krzywizny (NCM). Te punkty mogą być wykorzystane między innymi do rozpoznawania osób w sekwencjach wizyjnych, jak przedstawiono w [9]. Definicja minimum wklęsłości jest następująca: punkt NCM jest punktem leżącym na konturze pomiędzy punktami P_1 , P_2 (rys. 2) otoczki wypukłej, dla którego odległość od odcinka $[P_1P_2]$ jest największa. Punkty P_1 i P_2 są od siebie odpowiednio odległe, a sposób ich wyboru został dokładnie opisany w kolejnym punkcie pracy. Rysunek 2 przedstawia przykład punktu NCM.



Rys. 2. Punkt minimum wklęsłości pomiędzy punktami należącymi do otoczki (P_1, P_2).

Dodatkowo, oprócz minimum wklęsłości, także wypukłość konturu może być wykorzystana do opisu krzywizny. Metoda wykorzystująca wypukłość konturu nazywana jest punktami maksimum krzywizny (PCM). Tym razem, skrajne punkty P_1 i P_2 (rys. 2) wybierane są z otoczki wypukłej tak, że P_1 jest punktem zamykającym i -tą wklęsłość, a P_2 jest punktem poprzedzającym $i+1$ -ą wklęsłość. Przykład punktu PCM przedstawiono na rysunku 3. Punkt PCM leżący na konturze jest wybierany spośród par punktów P_1 i P_2 , tak, by jego odległość od odcinka $[P_1P_2]$ była możliwie jak największa.



Rys. 3. Punkt maksimum krzywizny P (czarny).

3.1 Wybór punktów

Algorytm wyznaczający punkty NCM zaczyna się krokiem S1 (rys.4) wyborem pary następujących po sobie punktów h_i, h_{i+1} z otoczki wypukłej. Jeśli przeanalizowany został już cały wektor otoczki, algorytm przechodzi od S2 do S9. Jeśli nie, to w kroku S3 obliczana jest długość odcinka „ a ” pomiędzy punktami h_i, h_{i+1} . Następnie w kroku S4 weryfikowane jest czy „ a ” jest większe od ustalonego progu. Jeśli długość odcinka „ a ” jest większa od wymaganego progu algorytm przystępuje do kroku S5, zaś w przeciwnym wypadku powraca do kroku S1 by wybrać kolejną parę punktów. W kroku S5, wybierany jest taki punkt C z wektora konturu, który znajduje się pomiędzy punktami h_i i h_{i+1} oraz jego odległość h od odcinka „ a ” jest największa.

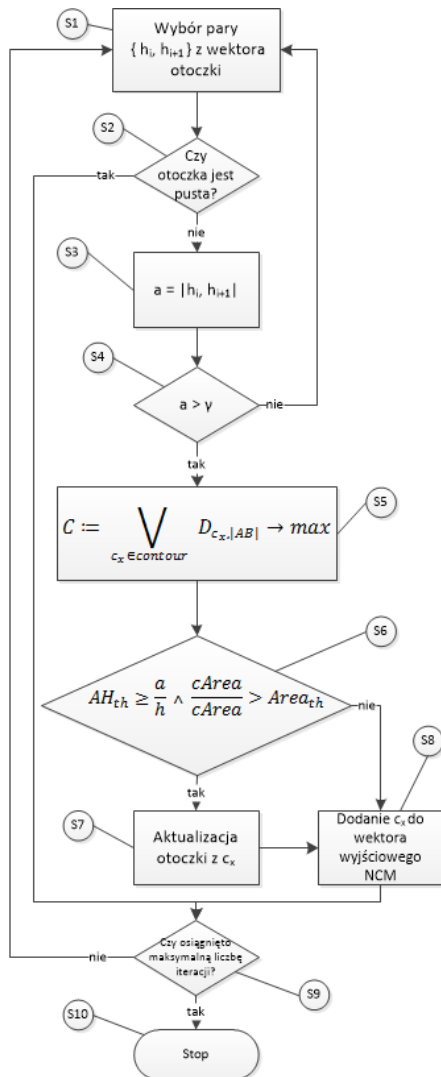
W kroku S6 sprawdzany jest warunek aktualizacji otoczki (ang. *UpdateHull*), aby dodać punkt C do otoczki wypukłej. Parametry warunku są następujące: AH_{th} jest to próg głębokości wklęsłości, $cncArea$ jest to pole wklęsłości pomiędzy punktami h_i oraz h_{i+1} , $cntArea$ jest powierzchnią całkowitą badanej postaci, natomiast $Area_{th}$ jest progiem minimalnym określającym stosunek powierzchni wklęsłości do powierzchni całego obiektu.

Jeśli warunek jest spełniony algorytm przystępuje do kroku S7 gdzie następuje aktualizacja otoczki, a następnie przechodzi do kroku S8 aby dodać punkt C do wyjściowego wektora NCM.

Na koniec, w kroku S9, sprawdzane jest czy osiągnięto wymaganą liczbę iteracji. Jeśli nie, to następuje

powrót do kroku S1 by rozpocząć wybieranie par punktów od początku.

Wspomniana w tym algorytmie metoda aktualizująca otoczkę wypukłą – *UpdateHull* jest autorską implementacją powszechnie dostępnego algorytmu znanego pod nazwą „Gift wrapping” lub „Jarvis march”. Ma ona za zadanie włączyć do otoczki wypukłej wyznaczony wcześniej punkt NCM tak, aby zachowana została jej definicja. Podczas wykonywania się tej metody do otoczki dodawana jest zatem minimalna ilość punktów z konturu tak, aby wektor *hull* zachował swoją ciągłość.



Rys. 4 Algorytm wyznaczania punktów NCM

Drugą częścią deskryptora opisywanego w tym punkcie są punkty PCM, które wyznaczane są podobnie do punktów NCM. Algorytm działa również na parach punktów z otoczki wypukłej, a jego przebieg wygląda następująco:

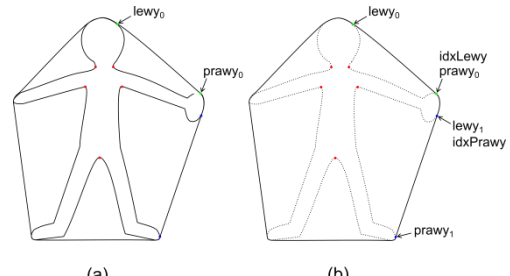
1. Wybieraj pary punktów $\{left_i, right_i\}$ tak długo, aż trafisz na parę spełniającą warunek $D_{|left_i, right_i|} > \delta$ uzyskując $\{left_0, right_0\}$ pokazaną na rysunku 5(a). Zapamiętaj indeks punktu $right_0$ w wektorze otoczki jako $idxLeft$.
2. Przesuwaj parę $\{left_i, right_i\}$, aż trafisz na następną parę spełniającą warunek NCM uży-

skując $\{left_1, right_1\}$ tak jak na rysunku 5(b). Zapamiętaj indeks punktu $left_1$ jako $idxRight$.

3. Z wektora otoczki wypukłej *hull*, pomiędzy indeksami $idxLeft$ i $idxRight$, wybierz taki punkt K_0 , dla którego odległość h_0 od odcinka $|right_0 left_1|$ jest największa.
4. Ustaw $idxLeft = idx(right_1)$ i kontynuuj postępowanie od punktu 2.

Podsumowując, kolejne punkty K_i wyliczane są analogicznie, maksymalizując odpowiadające im odległości h_i od odcinków $|right_i left_{i+1}|$. Algorytm wykonuje ostatnią iterację, gdy trafi na $left_n = left_0$. Rysunek 5 przedstawia naprzemienne operacje mające na celu wyznaczenie kolejnych par $idxLeft$ i $idxRight$, gdzie kolorem niebieskim zaznaczone zostały wyliczone w rezultacie punkty PCM.

Wektor wyznaczonych w tym algorytmie punktów K_i dodawany jest do wcześniej wyliczonych punktów NCM tworząc w ten sposób deskryptor sylwetki.



Rys. 5. Przykład wyboru punktów PCM.

4. WYNIKI EKSPERYMENTÓW

Badania przeprowadzone zostały na dwóch sekwencjach wizyjnych SD oraz jednej sekwencji HD, które zarejestrowane zostały w różnych lokalizacjach oraz w różnych warunkach pogodowych. Wszystkie eksperymenty przeprowadzono w obszarach aktywnych oraz podzielone były na dwie części. Pierwsza część dotyczyła rozpoznawania człowieka, gdyż konieczne było ustalenie, które z zarejestrowanych obiektów są ludźmi (krok „Detekcja poruszających się sylwetek” na rys. 1.) przed przystąpieniem do analizy zachowań. Druga część dotyczyła detekcji oraz analizy zachowań wykrytych w scenie ludzi.

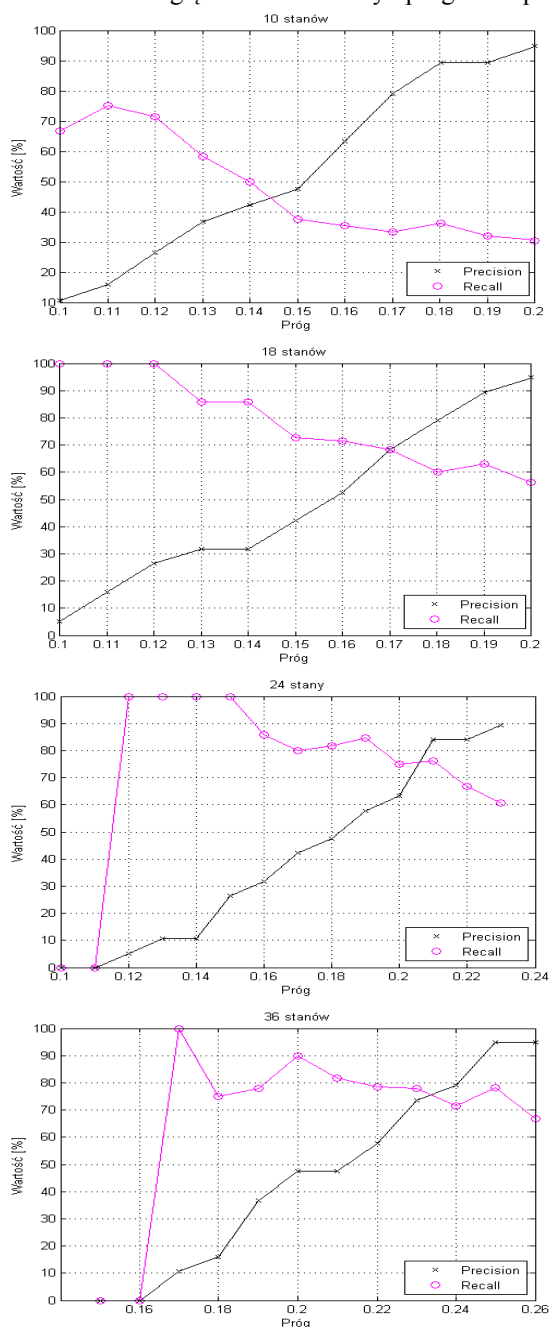
Autorzy przeanalizowali kilka metod klasyfikacji, z których większość należy do grupy algorytmów uczenia maszynowego. Ze względu na fakt, iż algorytmy te potrzebują przygotowanego zestawu treningowego przyjęto, że zbiór danych powinien być jak najmniejszy oraz jak najefektywniejszy. W związku z tym, przygotowany został zbiór danych treningowych składający się z 396 pozytywnych oraz 742 negatywnych obiektów. Badana była skuteczność klasyfikacji, zaś do jej oceny wykorzystano miary skuteczności znane w literaturze jako *precision* oraz *recall*, zdefiniowane następująco:

$$precision = \frac{TP}{TP+FP}, \quad recall = \frac{TP}{TP+FN},$$

gdzie TP jest zbiorem prawdziwie pozytywnych, FP jest zbiorem fałszywie pozytywnych, zaś FN jest zbiorem fałszywie negatywnych. Zgodnie z osiągniętymi wynikami (97,89% *precision* oraz 74,27% *recall*) do

dalszych eksperymentów wybrany został klasyfikator SVM z deskryptorami HOG.

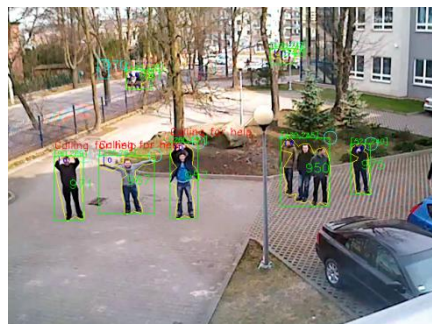
Skuteczność zaproponowanego rozwiązania sprawdzana była na podstawie czterech deskryptorów, które składały się z różnej liczby stanów (10, 18, 24 oraz 36 stanów). Każdy z nich sprawdzany był pod kątem skuteczności ze względu na zmienny próg akceptacji.



Rys. 6. Wyniki badań skuteczności działania systemu detekcji ludzkich zachowań (dla 10,18,24 i 36 stanów).

Najlepsze wyniki zostały uzyskane dla deskryptorów o 36 stanach, które osiągnęły 94,74% *precision* oraz 78,26% *recall* podczas gdy inny obiecujący deskryptor uzyskał niemal 90% *precision* kosztem utraty skuteczności detekcji (60,71% *recall*). Wyniki pokazały również, że deskryptory o większej liczbie stanów wymagają wyższego stopnia tolerancji by osiągnąć dobre wyniki detekcji. Ponadto zaobserwowano, że rozłożenie punktów na konturze sylwetki ma znaczny wpływ na jakość detekcji. Niektóre konfiguracje charakteryzowały się

lepszymi wynikami od innych. Wyniki detekcji przedstawione zostały na rysunkach 6 oraz 7.



Rys. 7. Wizualizacja wyników detekcji.

5. WNIOSKI

W niniejszym artykule przedstawiony został nowatorski system detekcji ludzkich zachowań w sekwencjach wizyjnych układów jednokamerowych. Zaproponowany system jest złożonym rozwiązaniem wykorzystującym kilka technik, które mają za zadanie wykryć poruszające się obiekty, śledzić je oraz rozpoznawać ich zachowania. Do klasyfikacji system ten wykorzystuje kombinację złożoną z punktów NCM oraz PCM wybieranych z konturu sylwetki. Wyniki eksperymentów pokazują, że zaproponowane rozwiązanie osiąga wysoki stopień detekcji równocześnie zachowując wysoki stopień precyzji. Rozwiązanie jest także odporne na warunki pogodowe i różnorodne realizacje zachowań. Co więcej, wyniki eksperymentów wykazały, że wpływ rozłożenia punktów na konturze zasługuje na dalsze badania.

PODZIĘKOWANIA

Projekt został sfinansowany ze środków Narodowego Centrum Nauki. Umowa nr 4775/B/T02/2011/40.

SPIS LITERATURY

- [1] H. Buxton and S. Gong, "Visual surveillance in a dynamic and uncertain world", *Artificial Intelligence*, 78(1-2):431–459, 1995.
- [2] S. S. Intille and A. F. Bobick. "A framework for recognizing multi-agent action from visual evidence", *AAAI Conference on Artificial intelligence*, pages 518–525, 1999.
- [3] Y. Du, F. Chen, W. Xu, and Y. Li., "Recognizing interaction activities using dynamic Bayesian network", *International Conference on Pattern Recognition*, pages 618–621, 2006.
- [4] T. Duong, H. Bui, D. Phung, and S. Venkatesh. "Activity recognition and abnormality detection with the switching hidden semi-Markov model", *IEEE Conference on Computer Vision and Pattern Recognition*, pages 838–845, 2005.
- [6] D. Kuettel, M. D. Breitenstein, L. V. Gool, and V. Ferrari, "What's going on? Discovering spatio-temporal dependencies in dynamic scenes", *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1951–1958, 2010.
- [7] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities - a survey", *IEEE Transactions on Circuits and Systems for Video Technology*, 18(11):1473–1488, 2008.
- [8] Y. Shi, A. Bobick, and I. Essa, "Learning temporal sequence model from partially labeled data", *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1631–1638, 2006.
- [9] L. Zhao; „Dressed Human Modeling, Detection, and Parts Localization”, The Robotics Institute, Carnegie Mellon University, Pittsburgh, 2001.