

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



Łódź, 8-10 kwietnia 2015

ROZPOZNAWANIE ZACHOWAŃ OSÓB Z WYKORZYSTANIEM PUNKTÓW ZNAJDUJĄCYCH SIĘ W ŚĄSIEDZTWIE PUNKTÓW EKSTREMUM KRZYWIZNY KONTURU CIAŁA

Streszczenie: W artykule przedstawiono metodę automatycznego rozpoznawania zachowań osób z wykorzystaniem modelu opartego o graf ukierunkowany. W zaproponowanym rozwiązaniu, jako punkty charakterystyczne, zastosowano punkty GDDP na konturze sylwetki, czyli punkty, które znajdują się w sąsiedztwie punktów minimum wklęsłości i maksimum wypukłości konturu sylwetki. Wyniki eksperymentów potwierdzają poprawę skuteczności rozpoznawania zachowań w stosunku do metody bazującej na punktach ekstremum krzywizny konturu sylwetki.

1. WSTĘP

W ostatnich latach znacząco wzrosła liczba kamer rejestrujących zachowania osób w różnych dziedzinach życia i jednocześnie spowodowała znaczny przyrost ilości materiału wizyjnego. W związku z tym wzrasta również zapotrzebowanie na aplikacje służące do opisu zawartości treści sekwencji wizyjnych, czy też streszczania sekwencji opartego na rozpoznawaniu zarejestrowanych zdarzeń. Analiza zachowań ludzi w sekwencjach wizyjnych zyskała na znaczeniu wraz z rozwojem możliwości wykorzystania obrazu wizyjnego w systemach ochrony i dozoru wizyjnego. Szczególnie znaczące jest zastosowanie takich rozwiązań w zautomatyzowanych systemach obserwacji obszarów publicznych i wykrywania niebezpiecznych zdarzeń. Tego rodzaju zadanie obserwacji jest bardzo wymagające dla operatora systemu jakim jest człowiek, ponieważ wymaga dużego skupienia przez dłuższy 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.

Zakres przeglądu został ograniczony do powszechnie stosowanych modeli grafowych wykorzystywanych w złożonym modelowaniu aktywności człowieka w zatłoczonych przestrzeniach publicznych. Istnieje wiele różnorodnych metod do modelowania zachowań w systemach jednokamerowych, do których należą m.in. modele oparte o grafy prawdopodobieństwa (np. *sieci Bayesa* [3, 4], *dynamiczne sieci Bayesa* [5-7], *sieci propagacji* [8]), modele semantyczne (np. *probabilistyczny model analizy semantycznej* [9], *model alokacji ukrytej zmiennej Dirichleta* [10, 11], *model hierarchicznego procesu Dirichleta* [12, 13]), *sieci Petriego* [14], podejścia syntaktyczne [15, 16] lub regułowe [17]. Ze względu na fakt, że zaproponowane rozwiązanie przynależy do

klasy metod opartych o sieci Bayesa, stąd ich bardziej szczegółowa dalsza analiza.

Sieci Bayesa są popularnym narzędziem do modelowania zachowań ze względu na duże możliwości reprezentacji i rozumienia niepewnych zdarzeń zarejestrowanych przez kamerę, a także możliwości obliczeniowych. Sieć Bayesa lub algorytm propagacji wierzeń jest acyklicznym grafem kierunkowym z węzłami reprezentującymi 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. Identyfikacja stanu w sieci Bayesa związana jest z analizą opisu tego stanu za pomocą wyznaczonych cech charakterystycznych w obrazie. W niniejszym artykule są to punkty o charakterystycznym rozkładzie na konturze.

Artykuł podzielony jest na 5 sekcji. W sekcji 2 przedstawiono poprzednie prace dotyczące systemów rozpoznawania zachowań osób oraz wyjaśniono szczegóły dotyczące punktów minimum wklęsłości (ang. Negative Curvature Minima, NCM) i maksimum wypukłości (ang. Positive Curvature Maxima, PCM) konturu sylwetki oraz punktów o równomiernym rozłożeniu (ang. Evenly Distributed Points, EDP), które zostały zaproponowane i przebadane przez autorów. W sekcji 3 przedstawiono szczegóły nowego rozwiązania proponowanego przez autorów dotyczącego punktów charakterystycznych GDDP (ang. Gaussian Density Distributed Points), które są głównym tematem niniejszego artykułu. Sekcja 4 przedstawia podsumowanie eksperymentów i osiągnięte rezultaty dla rozpoznawania przykładowych zachowań. Sekcja 5 zawiera podsumowanie i sugestie dotyczące dalszych prac.

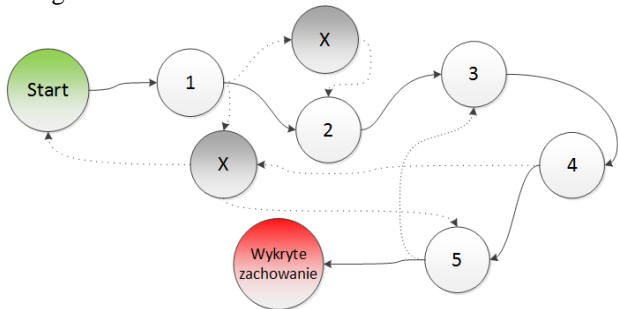
2. POPRZEDNIE PRACE

Rozważmy system oparty na pojedynczej, stacjonarnej kamerze, która obserwuje scenę. Zdarzenia występujące w scenie można interpretować (na poziomie ramek) jako pewną sekwencję stanów w czasie (rys. 1). Takie stany mogą być opisane z wykorzystaniem cech charakterystycznych obiektu. Do najbardziej popularnych technik detekcji i opisu cech charakterystycznych należą SIFT (ang. Scale Invariant Feature Transform), cechy Haar-podobne oraz HOG (ang. Histogram of Oriented Gradients)[19]. Niemniej jednak wyniki poka-

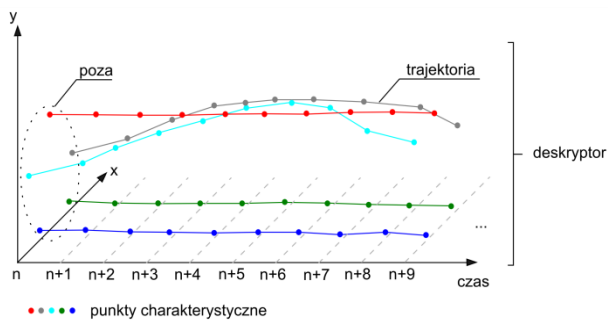
zane w kolejnym paragrafie dowodzą, że metoda opisana w tym artykule daje porównywalnie dobre wyniki jak metody wymienione powyżej, ale jest od nich mniej złożona obliczeniowo, ponieważ wymagana jest dużo mniejsza liczba cech charakterystycznych niż w pozostałych metodach.

Rozpoznawanie zachowania osób oparte jest na idei śledzenia punktów charakterystycznych [18]. Punkty te wybierane są z konturu wykrytego wcześniej człowieka i opisują jego pozę (rys. 2). Następnie, pozycje punktów są śledzone w czasie tworząc trajektorię (deskrytor). W kolejnym kroku trajektorie punktów charakterystycznych są porównywane z predefiniowanymi trajektoriami odpowiadającymi konkretnemu zachowaniu. Porównanie odbywa się poprzez wyznaczenie odległości euklidesowej dla pary odpowiadających sobie punktów. Zachowanie jest rozpoznane, jeżeli trajektoria punktów charakterystycznych nie różni się od trajektorii predefiniowanej o więcej niż wartość progowa.

Zbiór punktów definiujący pozę posiada różną konfigurację dla różnych zachowań. Innymi słowy, wygenerowana trajektoria punktów dla jednego typu zachowania jest inna niż trajektoria punktów wygenerowana dla innego zachowania.



Rys. 1 Zachowanie człowieka jako zbiór póz (stanów), Stany X oznaczają stany ukryte.



Rys. 2 Przykład zachowania człowieka jako zbiór póz. Punkty charakterystyczne tworzą trajektorię w czasie. Trajektoria uzyskana z zachowania "wzywanie pomocy".

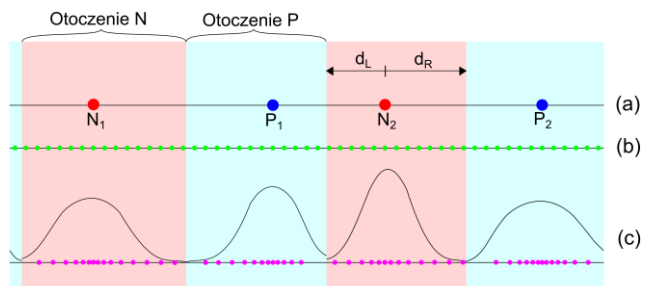
Istnieją dwie różne metody wyboru punktów charakterystycznych. Pierwszą z nich nazwano metodą punktów równomiernie rozłożonych [1] (ang. Evenly Distributed Points, EDP). Metoda ta polega na wyborze z góry założonej liczby punktów równomiernie rozłożonych na konturze sylwetki. To podejście pozwala na stopniową selekcję poziomu odwzorowania konturu.

W celu poprawy efektywności systemu wykorzystano inne metody selekcji punktów. Metody te noszą nazwę minimum wklęsłości (ang. Negative Curvature

Minima, NCM) oraz maksimum wypukłości (ang. Positive Curvature Maxima, PCM). W przeciwieństwie do metody EDP liczba punktów charakterystycznych nie jest stała i może się zmieniać w zależności od kształtu konturu człowieka. Zastosowanie metod NCM i PCM w rozpoznawaniu zachowań ludzi zostało opisane w [2].

3. PROPONOWANA METODA WYBORU PUNKTÓW CHARAKTERYSTYCZNYCH

W zaproponowanej metodzie punkty charakterystyczne są rozmieszczane zgodnie z rozkładem normalnym wokół punktów ekstremum krzywizny konturu (NCM i PCM). W tym celu wykorzystywane są znane już punkty równomiernie rozłożone na konturze (EDP), które są przesuwane w pobliże punktów ekstremum krzywizny tak, aby uzyskana gęstość rozmieszczenia punktów miała rozkład Gaussa. Tak uzyskane punkty charakterystyczne nazywamy w skrócie GDDP (ang. Gaussian Density Distributed Points). Na rys. 3 przedstawiona została ilustracja procesu wyznaczania punktów GDDP. Czerwone i niebieskie punkty oznaczają punkty NCM i PCM (rys. 3(a)). Zielone punkty na rys. 3(b) to punkty EDP, które po procesie zagęszczenia do rozkładu Gaussa są zgrupowane wokół punktów NCM i PCM. Sytuacja ta została przedstawiona na rys. 3(c).



Rys. 3 Wizualizacja procesu wyznaczania punktów rozłożonych zgodnie z rozkładem Gaussa (ang. Gaussian Density Distributed Points, GDDP).

Formalnie algorytm wyznaczania punktów GDDP w każdej ramce sekwencji wizyjnej przebiega następująco:

- Wyznacz zestawy punktów: NCM, PCM oraz EDP na konturze człowieka.
- Wyznacz wartość funkcji $h(x)$ zagęszczającej punkty EDP:

$$\begin{cases} g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{\sigma^2}} \\ \sigma = \frac{d_L+d_R}{4} \end{cases} \quad (1)$$

$$h(x) = \frac{1}{\sigma\sqrt{2\pi}} \left(1 - e^{-\frac{(x-\mu)^2}{\sigma^2}} \right) \quad (2)$$

gdzie μ jest pozycją rozpatrywanego punktu ekstremum krzywizny (na rys. 3 jest to punkt N_2), natomiast x jest indeksem punktu EDP. Odchylenie standardowe zostało dobrane proporcjonalnie do szerokości rozważanego przedziału.

- Wyznacz przesunięcie dla punktów EDP zgodnie z wzorem 3:

$$\begin{cases} h_d(x) = \frac{\mu-x}{|\mu-x|} \frac{d_L}{2} \frac{h(x)}{\sigma\sqrt{2\pi}}, & x < \mu \\ h_d(x) = \frac{\mu-x}{|\mu-x|} \frac{d_R}{2} \frac{h(x)}{\sigma\sqrt{2\pi}}, & x > \mu \end{cases} \quad (3)$$

gdzie d_L , d_R są połowami odległości do sąsiednich punktów ekstremum krzywizny (na rys. 3 są to punkty P_1 i P_2). Dzielenie przez $\sigma\sqrt{2\pi}$ w wzorze 3 odpowiada za normalizację funkcji zagęszczającej.

4. WYNIKI BADAŃ

Do badań wykorzystano sekwencje o rozdzielczości 1240x720, które zostały podzielone w następujący sposób:

- zbiór uczący – 9 osób,
- zbiór treningowy – 8 osób,
- zbiór testowy – 12 osób¹.

Autorzy sprawdzili popularne zbiory sekwencji testowych, które okazały się nie spełniać postawionych wymagań. W niniejszym artykule przedstawiono wyniki dotyczące zachowania *wzywaniu pomocy*. Oprogramowanie testowe przygotowano w języku C++.

Badana była skuteczność rozpoznawania zachowania, zaś do oceny algorytmów klasyfikacji wykorzystano popularne metryki *precision* oraz *recall*, które zdefiniowane są następująco:

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$recall = \frac{TP}{TP + FN} \quad (5)$$

gdzie TP oznacza zbiór prawdziwie pozytywny (ang. True Positive), FP jest zbiorem fałszywie pozytywnym (ang. False Positive) natomiast FN jest zbiorem fałszywie negatywnym (ang. False Negative).

Efektywność zaproponowanego rozwiązania została sprawdzona przy pomocy trzech deskryptorów, które składały się z różnej liczby stanów (10, 18 i 36 stanów). Badania przeprowadzono dla metod bazujących na wyborze różnych punktów kluczowych: NCM i PCM, EDP oraz GDDP. Wyniki przedstawione zostały w tabeli 1.

Tab. 1. Uzyskane wyniki

Metoda	Liczba stanów	Precision	Recall
EDP (40 pkt)	10	58,11%	100,00%
	18	47,93%	100,00%
	36	42,13%	100,00%
NCM i PCM	10	37,42%	89,71%
	18	63,25%	90,85%
	36	79,54%	95,68%
GDDP (40 pkt)	10	71,47%	97,33%
	18	82,63%	96,05%
	36	81,16%	98,97%

¹ Sekwencje testowe wykorzystane do badań są dostępne pod adresem: <http://www.multimedia.edu.pl/human-behavior>



Rys. 4 Górny obraz pokazuje rozłożenie punktów EDP (zielone), punktów NCM (czerwone), punktów PCM (niebieskie). Dolny obraz przedstawia rozłożenie punktów GDDP (magenta).

Zaproponowana metoda osiągnęła wyższą wartość *recall* względem metod bazujących na punktach NCM i PCM, ale mniejszą niż ta bazująca na punktach EDP. Uzyskano jednak znaczną poprawę parametru *precision* względem metod bazujących na punktach NCM i PCM, czy też EDP. Na podstawie przeprowadzonych badań można wywnioskować, że wpływ na skuteczność detekcji zachowań ma zarówno liczba punktów charaktery-

stycznych, jak i ich położenie. Szczególnie dobrze ilustruje to przykład punktów EDP, których przesunięcie w kierunku ekstremów krzywizny znacznie zmniejsza liczbę fałszywych detekcji.

Ponadto, skuteczność detekcji można również poprawić wykorzystując większą liczbę stanów systemu klasyfikującego badane zachowanie.

5. WNIOSKI

W zaproponowanym rozwiązaniu zastosowano jako punkty charakterystyczne punkty GDDP na konturze sylwetki, czyli punkty, które znajdują się w sąsiedztwie punktów minimum wklęsłości i maksimum wypukłości sylwetki. Nowa metoda została porównana z metodami bazującymi na punktach EDP oraz NCM i PCM. Wyniki eksperymentów pokazują, że zaproponowane rozwiązanie osiąga wyższy stopień skuteczności.

Detekcja zachowań jest na tyle złożonym zagadnieniem, że problematyczne jest sformułowanie pojedynczej metody, która będzie skuteczna niezależnie od rodzaju zachowania. Podobny problem występuje w przypadku różnego ustawienia obiektu względem kamery. Tak więc relacja pomiędzy poszczególnymi zachowaniami oraz metodami ich opisu wymaga dalszych badań. Najbardziej pożądanym byłby mechanizm automatycznego wyboru odpowiedniej metody opisu zachowania w zależności od przedstawionych wyżej warunków.

PODZIĘKOWANIA

Praca finansowana ze środków przyznanych przez Ministerstwo Nauki i Szkolnictwa Wyższego na działalność statutową w roku 2014 w ramach zadania pn. „Teoria i algorytmy wielowymiarowego przetwarzania sygnałów”.

SPIS LITERATURY

- [1] Ł. Kamiński, K. Kowalak, P. Gardziński, S. Maćkowiak, "Human Activity Interpretation Using Evenly Distributed Points on the Human Hull", *Lecture Notes in Computer Science, Computer Vision and Graphics*, vol. 8671, str. 316-323, Springer-Verlag, 2014.
- [2] K. Kowalak, Ł. Kamiński, P. Gardziński, S. Maćkowiak, R. Hofman, "Human Behavior Recognition Using Negative Curvature Minima and Positive Curvature Maxima Points", *New Research in Multimedia and Internet Systems, Advances in Intelligent Systems and Computing*, vol. 314, str. 57-66, Springer, 2015
- [3] H. Buxton and S. Gong, "Visual surveillance in a dynamic and uncertain world", *Artificial Intelligence*, 78(1-2):431-459, 1995.
- [4] S. S. Intille and A. F. Bobick, "A framework for recognizing multi-agent action from visual evidence", *AAAI Conference on Artificial intelligence*, str. 518-525, 1999.
- [5] Y. Du, F. Chen, W. Xu, and Y. Li., "Recognizing interaction activities using dynamic Bayesian network", *International Conference on Pattern Recognition*, str. 618-621, 2006.
- [6] 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*, str. 838-845, 2005.
- [7] S. Gong and T. Xiang, "Recognition of group activities using dynamic probabilistic networks", *IEEE International Conference on Computer Vision*, str. 742-749, 2003.
- [8] Y. Shi, A. Bobick, and I. Essa, "Learning temporal sequence model from partially labeled data", *IEEE Conference on Computer Vision and Pattern Recognition*, str. 1631-1638, 2006.
- [9] J. Li, S. Gong, and T. Xiang, "Global behavior inference using probabilistic latent semantic analysis", *British Machine Vision Conference*, str. 193-202, 2008.
- [10] T. Hospedales, S. Gong, and T. Xiang, "A Markov clustering topic model for mining behavior in video", *IEEE International Conference on Computer Vision*, str. 1165-1172, 2009.
- [11] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model", *IEEE Conference on Computer Vision and Pattern Recognition*, str. 935-942, 2009.
- [12] 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*, str. 1951-1958, 2010.
- [13] X. Wang, X. Ma, and W. E. L. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):539-555, 2009.
- [14] M. Albanese, R. Chellappa, V. Moscato, A. Picariello, V. S. Subrahmanian, P. Turaga, and O. Udrea, "A constrained probabilistic Petri net framework for human activity detection in video", *IEEE Transactions on Multimedia*, 10(6):982-996, 2008.
- [15] M. Brand, "Understanding manipulation in video", *International Conference on Automatic Face and Gesture Recognition*, pages 94-99, 1996.
- [16] Y. A. Ivanov and A. F. Bobick, "Recognition of visual activities and interactions by stochastic parsing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):852-872, 2000.
- [17] H. Dee and D. Hogg, "Detecting inexplicable behavior", *British Machine Vision Conference*, str. 477-486, 2004.
- [18] S. Maćkowiak, P. Gardziński, Ł. Kamiński, K. Kowalak, "Human Activity Recognition in Multiview Video", *11th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, str. 148-153, South Korea, 2014
- [19] N. Dalal; B. Triggs; , "Histograms of Oriented Gradients for Human Detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Montbonnot, France, 2005.