

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



Kraków, 27-29 czerwca 2016

## NIENADZOROWANA DETEKcja ANORMALNYCH ZACHOWAŃ TŁUMU W DOZOROWYCH SEKWENCJACH WIZYJNYCH

### UNSUPERVISED ABNORMAL CROWD ACTIVITY DETECTION IN SURVEILLANCE SYSTEMS

**Streszczenie:** W niniejszym artykule opisano nienadzorowaną metodę detekcji anormalnych zachowań tłumu w dozorowych sekwencjach wizyjnych. Proponowane rozwiązanie wykorzystuje deskryptory standardu MPEG-7 do opisu sceny oraz algorytm Particle Filter do klasyfikacji. Badania przeprowadzono na ogólnodostępnej bazie sekwencji testowych UMN. Otrzymane wyniki są porównywalne do wyników uzyskiwanych przez metody nadzorowane.

**Abstract:** We propose an unsupervised method for abnormal crowd activity detection in surveillance systems. Proposed solution is using MPEG-7 Motion Activity descriptors and Particle Filter algorithm for classification. The experiments were performed on UMN dataset sequences. The detection results are comparable to results obtained by supervised methods.

**Słowa kluczowe:** filtr cząsteczek, nienadzorowana detekcja anomalii, UMN

**Keywords:** particle filter, UMN, unsupervised anomaly detection

#### 1. WPROWADZENIE

W ostatnich latach obserwuje się wzrost liczby kamer w przestrzeni publicznej. W związku z tym rośnie również ilość materiału wizyjnego podlegającego analizie. Istotnym zagadnieniem, ze względów bezpieczeństwa, jest np. obserwacji tłumu osób, w którym szybka detekcja anormalnych zachowań staje się wyjątkowo krytyczna. Detekcja anormalnych zdarzeń w tłumie ludzi odgrywa ważną rolę w wielu zastosowaniach, np. dozór wizyjny czy kontrola bezpieczeństwa.

Głównym założeniem metod detekcji anormalnych zdarzeń jest znajomość wzorca zdarzeń normalnych. Wzorzec ten jest poznawany w procesie uczenia ze zbioru definiującego zdarzenia normalne. Następnie na etapie klasyfikacji aktualne dane są porównywane ze wzorcem i na tej podstawie stwierdza się, czy dane zdarzenie jest normalne czy nie.

Wymaganie znajomości gotowego wzorca zdarzeń normalnych znacząco komplikuje zastosowanie metod detekcji zdarzeń anormalnych w rzeczywistych

systemach dozorowych. Ponadto, uprzednio wyuczone, normalne zachowania w pewnych sytuacjach mogą stać się anormalne. Przykładem mogą być ruchome schody. Zakładając, że anormalny jest ruch ludzi pod prąd, przy każdorazowej zmianie kierunku ruchu schodów konieczne byłoby dostarczenie do systemu informacji z zewnątrz. W przeciwnym wypadku dochodziłoby do generowania fałszywych alarmów.

Zagadnienie detekcji anomalii w tłumie ludzi cieszy się dużym zainteresowaniem w ostatnich latach. Metody detekcji anomalii, ze względu na sposób uczenia, mogą zostać podzielone na 3 grupy.

Pierwsza grupa traktuje zagadnienie detekcji jako problem binarny. Są to metody w pełni nadzorowane. W celu wytrenowania klasyfikatora wymagane są nie tylko próbki zachowań normalnych, ale także próbki zachowań anormalnych. Przykładowo, algorytmy [3,6,10] wykorzystują klasyfikator SVM w celu detekcji anormalnych zachowań tłumu.

Szkolenie klasyfikatora jest złożonym zagadnieniem, ze względu na dużą liczbę możliwych anormalnych wzorców zachowań. Druga grupa algorytmów detekcji nie wymaga znajomości wzorców zdarzeń anormalnych. Wzorce zdarzeń normalnych są w dalszym ciągu wymagane. Do tej grupy algorytmów można zaliczyć [1,17,21].

Niniejsza praca należy do trzeciej grupy algorytmów. Należą do niej algorytmy nienadzorowane, czyli takie, gdzie informacja o wzorcach normalnych i anormalnych nie jest wymagana. Kilka algorytmów należących do tej grupy zostało już zaprezentowanych, ale ich zastosowanie ogranicza się głównie do niezatłoczonych scen [14,15,16,24]. Metoda prezentowana w niniejszym artykule znajduje zastosowanie zarówno w przypadku scen zatłoczonych jak i niezatłoczonych, gdzie wzorzec normalnego zachowania zmienia się w czasie.

Większość prac nad wykrywaniem anomalii skupia się na opracowaniu deskryptora charakteryzującego tłum ludzi np. social force model [5,23], czy mixture of dynamic textures [9]. Zamiast opracowywać nowy deskryptor, wykorzystano deskryptor Motion Activity [8,18] ze standardu MPEG-7. Deskryptor ten został już

wykorzystany przez [7], ale metoda ta jest nadzorowana. Natomiast, jako klasyfikator wykorzystany został Particle Filter (PF) [12,20], który jest powszechnie znaną techniką, wykorzystywaną w śledzeniu ruchomych obiektów [2,19,22]. Według wiedzy autorów, jest to pierwsze wykorzystanie tej techniki w celu nienadzorowanej detekcji anomalii w tłumie ludzi.

Artykuł podzielony jest na 5 sekcji. W sekcji 2 przedstawiono sposób wyznaczenia deskryptora Motion Activity. W sekcji 3 znajdują się szczegółowe informacje na temat równań aktualizacji i predykcji algorytmu PF. Badania proponowanego rozwiązania znajdują się w sekcji 4, wraz z omówieniem wyników. W sekcji 5 zawarto podsumowanie i sugestie dotyczące dalszych prac.

## 2. DESKRYPTORY RUCHU

W proponowanym rozwiązaniu zdecydowano się na wykorzystanie deskryptorów obecnych w standardzie MPEG-7 [18]. Deskryptory te należą do grupy deskryptorów Motion Activity [8]. Ich obliczenie opiera się na prostych operacjach na składowych wektorów ruchu [11]. W związku z tym ich wyznaczenie może odbywać się w obrazie skompresowanym, co znacznie przyspiesza proces ekstrakcji cech.

Pierwszym z deskryptorów jest intensywność ruchu. Służy on do określenia dynamiki tłumy. Im wyższa jest jego wartość, tym większy jest ruch w scenie. Jego wyznaczenie polega na zsumowaniu energii wszystkich wektorów ruchu w ramce obrazu. Zakładając, że energia  $i$ -tego wektora ruchu jest równa:

$$E_i = x_i^2 + y_i^2, \quad (1)$$

gdzie:  $x$  – składowa pozioma wektora ruchu,  $y$  – składowa pozioma wektora ruchu, intensywność ruchu wynosi:

$$E_c = \sum_{i=0}^N E_i. \quad (2)$$

Drugim deskryptorem jest 8 przedziałowy histogram kierunków ruchu. Deskryptor ten opisuje kierunki ruchu obiektów w scenie. W celu jego wyznaczenia obliczany jest kąt dla każdego wektora ruchu. W kolejnym kroku tworzony jest histogram, którego przedziały oddalone są od siebie co 45 stopni.

Trzeci deskryptor opisuje liczbę i rozmiar ruchomych obszarów w scenie. Deskryptor ten składa się z 3 wartości: NSR, NMR, NLR. Wartości te są wyznaczone z macierzy energii wektorów ruchu poddanej progowaniu. Macierz ta wyznaczana jest dla każdej ramki obrazu. Zakładając, że  $E_{avg}$  to średnia energia wektora ruchu w ramce obrazu, a jego współrzędne wynoszą  $(i, j)$ , macierz ta dana jest wzorem:

$$E_{i,j}^{thrs} = \begin{cases} E_{i,j} & \text{jeżeli } E_{i,j} > E_{avg} \\ 0 & \text{jeżeli } E_{i,j} \leq E_{avg} \end{cases}. \quad (3)$$

Wyznaczenie parametrów NSR, NMR, NLR polega na skanowaniu rastrowym macierzy  $E^{thrs}$  i zliczaniu długości ciągów niezerowych energii wektorów ruchu. Zakładając, że  $W$  oznacza szerokość macierzy  $E^{thrs}$ ,

a  $N$  długość bieżącego ciągu, to wartości te dane są następującą zależnością:

$$\begin{aligned} NSR &++ \text{ jeżeli } N > \frac{2W}{3} \\ NMR &++ \text{ jeżeli } \frac{W}{3} < N \leq \frac{2W}{3}. \\ NLR &++ \text{ jeżeli } N \leq \frac{W}{3} \end{aligned} \quad (4)$$

Czwarty deskryptor opisuje intensywność ruchu w największym ruchomym obszarze w stosunku do całego ruchu w scenie. Deskryptor ten, znajduje zastosowanie w przypadku skupiania się tłumy bądź jego rozchodzenia się. Wartość tego deskryptora opisana jest następującym wzorem:

$$E_{ratio} = \frac{E_{area}}{E_c}, \quad (5)$$

gdzie:  $E_{area}$  oznacza energię największego obszaru ruchomego.

Podsumowując, ostateczny deskryptor wyznaczony dla ramki obrazu opisany jest 13 elementowym wektorem:

$$descr = [d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, NSR, NMR, NLR, E_c, E_{ratio}], \quad (6)$$

gdzie:  $d_1$ - $d_8$  oznaczają przedziały histogramu kierunku ruchu.

## 3. ROZPOZNAWANIE ANOMALII

W celu wykrycia, które ramki obrazu zawierają anormalne zachowania tłumy, zdecydowano się na wykorzystanie algorytmu PF [12]. Algorytm ten jest implementacją rekursywnego filtru Bayesa poprzez symulację Monte Carlo. Jego celem jest estymacja funkcji gęstości prawdopodobieństwa na podstawie pomiarów. Zakładając, że tłum ludzi realizuje pewną sekwencję stanów, możemy założyć, że teraźniejszy stan  $x_k$ , zależy tylko od stanu poprzedniego:

$$x_k = f_k(x_{k-1}, v_{k-1}), \quad (7)$$

gdzie:  $f(.)$  jest pewną nieliniową funkcją, a  $v$  oznacza szum procesu. W takim przypadku, możliwa jest estymacja stanu  $x_k$  na podstawie pomiaru  $z_k$ :

$$z_k = h_k(x_k, n_k), \quad (8)$$

gdzie  $h(.)$  jest nieliniową funkcją, natomiast  $n$  oznacza szum pomiaru. Szczegółowe informacje na temat algorytmu PF można znaleźć w [20].

Algorytm PF wykorzystuje cząsteczki, które w analizowanym przypadku opisują stan, w którym przebywa tłum ludzi. Liczba cząsteczek w algorytmie jest stała i wynosi  $N = 100$ . W rozważanym przypadku cząsteczka jest 13 elementowym wektorem, co odpowiada rozmiarowi deskryptora.

Estymacja stanu tłumy odbywa się w dwóch następujących po sobie krokach: predykcji i aktualizacji. Na etapie predykcji wykorzystywany jest model przejścia, który ma na celu przewidzenie, w jakim stanie znajdzie się tłum. Uwzględniając losową naturę ruchu ludzi w tłumie zdecydowano się na wykorzystanie modelu chodu losowego, który polega na dodaniu do

składowych cząsteczki odpowiednio spreparowanego szumu:

$$x_k = Ax_{k-1} + Bv_{k-1}, \quad (9)$$

gdzie:  $A$ ,  $B$  oznaczają macierze jednostkowe, a  $v$  oznacza szum procesu.

W kolejnym etapie dostępny jest pomiar  $z_k$  i możliwe jest przeprowadzenie aktualizacji wag cząsteczek. W rozważanym przypadku pomiar  $z_k$  jest równy deskryptorowi  $descr$  wyznaczonemu dla aktualnie przetwarzanej ramki obrazu. Aktualizacja wag  $w_k$  odbywa się poprzez obliczenie wartości funkcji wykładniczej z ujemnej odległości Mahalanobisa pomiędzy cząsteczką  $x_k$ , a pomiarem  $z_k$ :

$$w_k = \exp(-MD(descr, x_k)). \quad (10)$$

Waga cząsteczki ma wymiar prawdopodobieństwa. Oznacza to, że im wyższa jest wartość wagi  $w_k$  dla cząsteczki  $x_k$ , tym bardziej prawdopodobne, że stan, w którym przebywa tłum jest normalny. Anomalia jest wykryta, gdy wagi wszystkich cząsteczek są mniejsze niż predefiniowana wartość progowa  $thrsh$ . Innymi słowy, anormalne zdarzenie jest wykryte, gdy obliczony deskryptor nie pasuje do przewidywanego stanu tłumy.

W celu uniknięcia degeneracji algorytmu PF [20], w końcowym etapie cząsteczki z niską wagą zastępowane są przez nowe cząsteczki, których waga jest wyższa. Pozwala to na ciągłe śledzenie zmieniającego się stanu tłumy.

#### 4. WYNIKI BADAŃ

Badania zostały przeprowadzone na bazie sekwencji UMN [25]. Baza ta, zawiera 11 sekwencji testowych. Każda z nich rozpoczyna się normalnym zachowaniem, przedstawiającym spacerujących ludzi, po którym następuje rozbiegnięcie się tłumy – zdarzenie anormalne. Przykładowe ramki obrazu z tej bazy danych zostały przedstawione na rysunku 1. W związku z tym, że cały tłum bierze udział w zdarzeniu anormalnym, baza ta została wykorzystana do ewaluacji efektywności algorytmu dla zdarzeń globalnych.

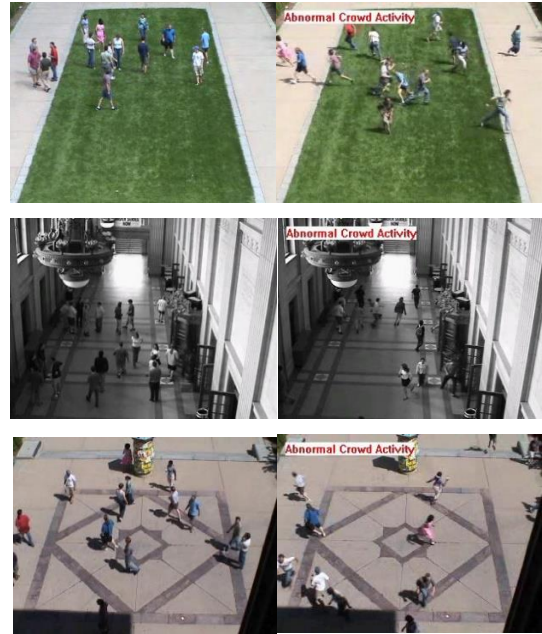
Twórcy bazy UMN ręcznie oznaczyli ramki obrazu przedstawiające anormalne zachowanie tłumy. Pozwala to na wykorzystanie popularnych metryk *precision* oraz *recall*, które zdefiniowane są następująco:

$$precision = \frac{TP}{TP+FP}, \quad (11)$$

$$recall = \frac{TP}{TP+FN}, \quad (12)$$

gdzie:  $TP$  oznaczana zbiór prawdziwe pozytywny (ang. TP – True Positive),  $FP$  jest zbiorem fałszywie pozytywnym (ang. False Positive), natomiast  $FN$  jest zbiorem fałszywie negatywnym (ang. False Negative). Wyznaczenie wartości *precision* oraz *recall* dla różnych wartości progu akceptacji  $thrsh$  pozwala na wykreślenie krzywej *ROC* (ang. Receiver Operating Characteristic). Następnie, w celu porównania badanej metody z innymi algorytmami wykorzystano miarę *AUC* (ang. Area Under the Curve), czyli obszar pod krzywą *ROC*. Im wyższa

wartość, tym lepsza ogólna sprawność algorytmu. Wyniki zostały przedstawione w tabeli 1.



Rys. 1. Przykładowe ramki obrazu z bazy UMN. W lewej kolumnie znajduje się przykład normalnego zachowania, a w prawej przykład zachowania anormalnego.

Tab. 1. Porównanie wyniku proponowanego algorytmu z innymi metodami.

Nazwa metody	AUC
Yang 2012 [3]	0.975
Mehran 2009 [5]	0.960
<b>Proponowany algorytm</b>	<b>0.945</b>
Zhao 2011 [13]	0.940
Wang 2012 [4]	0.900

Skuteczność przedstawionej metody jest porównywalna z innymi metodami. Warto zauważyć, że wszystkie te metody wymagają jednak próbek normalnego zachowania w celu wyuczenia klasyfikatora. Proponowana metoda jest pozbawiona tego wymagania. Model zachowania normalnego budowany jest na bieżąco w trakcie trwania sekwencji bez konieczności wcześniejszej inicjalizacji algorytmu. Uzyskany wynik potwierdza skuteczność algorytmu PF zastosowanego w celu detekcji anomalii w tłumie ludzi.

#### 5. PODSUMOWANIE

Niniejszy artykuł dotyczył nienadzorowanej detekcji anormalnych zachowań w tłumie ludzi. W zaproponowanym rozwiązaniu wykorzystano deskryptory MPEG-7 w celu scharakteryzowania ruchu występującego w obserwowanej scenie. W celu detekcji anormalnego zachowania tłumy wykorzystano algorytm PF. Według wiedzy autorów jest to pierwsze wykorzystanie tego algorytmu w celu detekcji anomalii w tłumie ludzi. W przeciwieństwie do innych metod prezentowany algorytm nie potrzebuje znajomości żadnych próbek zachowań normalnych i anormalnych.

Proponowana metoda jest w pełni nienadzorowana, co ułatwia jej praktyczną implementację, gdyż możliwa jest detekcja różnych typów anomalii. Badania zostały przeprowadzone na ogólnodostępnej bazie sekwencji testowych UMN. Uzyskane wyniki demonstrują wysoką efektywność w odniesieniu do innych metod, które wymagają uczenia.

Detekcja anomalii w tłumie ludzi jest złożonym zagadnieniem. Proponowana metoda jest w stanie rozpoznać obecność anomalii w scenie, natomiast nie jest w stanie wskazać jej lokalizacji. Jest to spowodowane wyznaczeniem deskryptora dla całej ramki obrazu. Następnym etapem prac nad algorytmem będzie lokalizacja anomalii w tłumie ludzi.

## PODZIĘKOWANIA

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

## LITERATURA

- [1] Wu S., Wong H.-S., Yu Z.. 2014. „A Bayesian model for crowd escape behavior detection”. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(1): 85-98.
- [2] Okuma K., Taleghani A., Freitas N.D., Little J.J., Lowe D.G.. 2004. „A boosted particle filter: multitarget detection and tracking”. *Proceedings of Eighth European Conference on Computer Vision*, 28-39.
- [3] Yang H., Cao Y., Wu S., Lin W., Zheng S., You Z.. 2012. „Abnormal crowd behavior detection based on local pressure model”. *Signal & Information Processing Association Annual Summit and Conference*, 1-4.
- [4] Wang B., Ye M., Li X., Zhao F., Ding J.. 2012. „Abnormal crowd behavior detection using high-frequency and spatio-temporal features”. *Machine Vision and Applications*, 23(3): 501-511.
- [5] Mehran R., Oyama A., Shah M.. 2009. „Abnormal crowd behavior detection using social force model”. *IEEE Conference on Computer Vision and Pattern Recognition*, 935-942.
- [6] Miao Y., Song J.. 2014. „Abnormal Event Detection Based on SVM in Video Surveillance”. *IEEE Workshop on Advanced Research and Technology in Industry Applications*, 1379-1383.
- [7] Liao H., Xiang J., Sun W., Feng Q., Dai J.. 2011. „An Abnormal Event Recognition in Crowd Scene”. *Sixth International Conference on Image and Graphics*, 731-736.
- [8] Divakaran Ajay. 2001. „An Overview of MPEG-7 Motion Descriptor and Their Applications”. *Computer Analysis of Image and Patterns, Lecture Notes in Computer Science*, 2124: 29-40.
- [9] Li W., Mahadevan V., Vasconcelos N.. 2014. „Anomaly Detection and Localization in Crowded Scenes”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(1): 18-32.
- [10] Lin H., Deng J.D., Woodford B.J.. 2015. „Anomaly detection in crowd scenes via online adaptive one-class support vector machines”. *IEEE International Conference on Image Processing*, 2434-2438.
- [11] ISO/IEC 14496-10. 2010. „Coding of Audio-Visual objects – Part 10: Advanced Video Coding”.
- [12] Isard M., Blake A.. 1998. „CONDENSATION – Conditional Density Propagation for Visual Tracking”. *International Journal of Computer Vision*, 29(1): 5-28.
- [13] Zhao J., Xu Y., Yang X., Yan Q.. 2011. „Crowd instability analysis using velocity-field based social force model”. *IEEE Visual Communications and Image Processing*, 1-4.
- [14] Ito Y., Kitani K.M., Bagnell J.A., Hebert M.. 2012. „Detecting Interesting Events Using Unsupervised Density Ratio Estimation”. *Computer Vision – ECCV 2012. Workshops and Demonstrations, Lecture Notes in Computer Science*, 7585: 151-161.
- [15] Boiman O., Irani M.. 2005. „Detecting irregularities in images and in video”. *Tenth IEEE International Conference on Computer Vision*, 1: 462-469.
- [16] Zhong H., Shi J., Visontai M.. 2004. „Detecting unusual activity in video”. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2: 819-826.
- [17] Breitenstein M.D., Grabner H., Van Gool L.. 2009. „Hunting Nessie – Real-time abnormality detection from webcams”. *IEEE 12th International Conference on Computer Vision Workshops*, 1243-1250.
- [18] ISO/IEC 15938-3:2002/Amd 3:2009. *Information Technology – Multimedia content description interface – Part 3: Visual, Amendment 3: Image signature tools*.
- [19] Khan Z., Balch T., Dellaert F.. 2005. „MCMC-based particle filtering for tracking a variable number of interacting targets”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(11): 1805-1819.
- [20] Doucet A., Godsill S., Andrieu C.. 2000. „On sequential Monte Carlo sampling methods for Bayesian filtering”. *Statistics and Computing*, 10(3): 197-208.
- [21] Zhao B., Fei-Fei L., Xing E.P.. 2011. „Online detection of unusual events in videos via dynamic sparse coding”. *2011 IEEE Conference on Computer Vision and Pattern Recognition*, 3313-3320.
- [22] Qi Z., Ting R., Husheng F., Jinlin Z.. 2012. „Particle Filter Object Tracking Based on Harris-SIFT Feature Matching”. *Procedia Engineering*, 29: 924-929.
- [23] Helbing D., Molnar P.. 1995. „Social Force Model for Pedestrian Dynamics”. *Physical review E*, 51(5): 4282-4286.
- [24] Wang X., Ma X., Grimson E.. 2007. „Unsupervised Activity Perception by Hierarchical Bayesian Models”. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1-8.
- [25] UMN. Unusual crowd activity dataset. <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>