

Automatyczne wykrywanie osób w nagraniach uzyskanych przy niedostatecznym oświetleniu

Streszczenie. Artykuł przedstawia system do detekcji osób na nagraniach pochodzących z monitoringu miejskiego na otwartej przestrzeni. Proponowany system został przetestowany w trudnych, nocnych warunkach oświetlenia. W celu polepszenia jakości zarejestrowanych sekwencji wideo zaproponowano algorytm lokalnej poprawy kontrastu. Dzięki niemu detekcja obiektów ruchomych za pomocą GMM (Gaussian mixture model) oraz analizy BLOB (binary large object) jest bardziej precyzyjna. Dodatkowo ruchome obiekty wykryte w obrazie binarnym są śledzone przy użyciu filtru Kalmana, co zwiększa skuteczność algorytmu wykrywającego osoby. W artykule omówiono również dobór parametrów programu oraz sposób akwizycji obrazów.

Abstract. The article presents the issue related to the intelligent analysis of video sequences, which are obtained from the city monitoring. Analysis of people detection, who passed under the camera in the outdoor scenes, has been tested in low lighting conditions (during the night). In order to improve the quality of acquired video sequences, local contrast enhancement algorithm was used. Thanks to this, detection of moving objects with the use of the GMM (Gaussian mixture model) and BLOB (binary large object) analysis is more precise. In addition, detected moving objects in the binary image are tracked with the use of Kalman filter, which increases the efficiency of people detection. Selection of algorithm parameters and video acquisition method were also discussed. (**Automatic detection of people from the recordings obtained in low lighting conditions**).

Słowa kluczowe: CCTV, inteligentna analiza wideo, poprawa jakości obrazu, lokalna poprawa kontrastu, analiza BLOB, detekcja i śledzenie obiektów ruchomych.

Keywords: CCTV, intelligent video analysis, improving image quality, local contrast enhancement, BLOB analysis, moving objects detection and tracking.

Wprowadzenie

Inteligentna analiza sekwencji wizyjnych staje się coraz bardziej popularna w systemach CCTV (closed circuit television) stosowanych w środowisku miejskim [1, 2]. Obserwacja tłumu bądź zliczanie pojedynczych osób przez uprawnione służby wpływa na poprawę bezpieczeństwa w obszarach zurbanizowanych. Wspomniane analizy są wykorzystywane podczas badań marketingowych. Stosowanie podanych metod powoduje, że zbyteczna staje się manualna analiza sekwencji wizyjnych.

Dotychczas systemy zajmujące się na przykład zliczaniem osób były stosowane wewnątrz budynków [3, 4] lub w miejscach dobrze oświetlonych [5–7]. Autorzy niniejszego artykułu pokonali te ograniczenia i zaproponowali algorytm pozwalający na prawidłowe wykrywanie osób nawet w niedostatecznych warunkach oświetleniowych. Algorytm przetestowano w nocy, w lokalizacji lekko oświetlonej przez uliczne latarnie. Dalszym etapem detekcji i śledzenia osób mogłoby być ich zliczanie.

Do akwizycji obrazów wykorzystano kolorową kamerę bez doświetlaczy na podczerwień. Zastosowanie kamery termowizyjnej lub umożliwiającej uzyskanie dodatkowej informacji o głębi (w celu uzyskania sceny 3D) byłoby bardziej kosztowne [8]. W warunkach bardzo słabego oświetlenia trudno zarejestrować obrazy nadające się do analizy przez system wykrywający obiekty. Spowodowane jest to: niewielką czułością matryc w kamerach, małym zakresem dynamicznym i wysokim poziomem zaszumienia uzyskanych obrazów, uniemożliwiającym proponowanemu algorytmowi odróżnienie postaci od tła. Czytelność obrazów w systemach CCTV można poprawiać za pomocą lokalnych zmian jasności obrazu [9–11]. Metoda ta może być stosowana wraz z technikami stereowizyjnymi zwiększającymi koncentrację obserwatora [12]. Szum impulsowy o zmiennych wartościach może być usuwany za pomocą metod iteracyjnych wykorzystujących dynamicznie dobierany próg detekcji [13, 14]. W niniejszym artykule autorzy zastosowali metodę lokalnej poprawy kontrastu obrazu, co pozytywnie wpłynęło na jakość detekcji obiektów ruchomych.

Istotną kwestią jest również rozróżnienie osób od innych obiektów znajdujących się w kadrze kamery (samochody,

zwierzęta). Algorytm powinien służyć wyłącznie do detekcji pieszych celem na przykład ich dalszego zliczania. Zastosowano analizę BLOB (z ang. *binary large object*) i następujące po niej operacje na obrazie binarnym. W celu rozróżnienia typu obiektu użyto współczynnika Fereta. Algorytm zaimplementowano w środowisku Matlab wykorzystując funkcje wspierające przetwarzanie obrazów i sekwencji wideo.

Akwizycja obrazów

Typowo kamery zapisują rejestrowany obraz w trybie wykorzystującym 256 poziomów kwantowania, a więc 8 bitów pamięci dla każdej składowej R, G i B pojedynczego piksela. Jednak ta liczba poziomów nie odzwierciedla całego zakresu dynamicznego matrycy. W związku z tym, w przyszłości, poprawę jakości obrazów zawierających sceny o niewielkim natężeniu światła (tam, gdzie np. ze względu na duże odległości nie można zastosować doświetlaczy na podczerwień), można uzyskać stosując m.in. czulsze matryce rejestrujące obraz z większą liczbą poziomów kwantowania pojedynczych pikseli. Opisane w literaturze metody powiększania zakresu dynamicznego obrazów mogą być stosowane wtedy, kiedy dostępne są dane w formacie RAW [15, 16]. W obecnej fazie rozwoju, większość systemów wizyjnych, w tym proponowany algorytm detekcji i śledzenia osób, oferuje przetwarzanie obrazów o 8-bitowej głębi.

Do akwizycji obrazów wykorzystano więc aparat Canon 70D. Wykonano zdjęcia seryjne, które zapisano w 14-bitowym formacie RAW. Pobierano 7 klatek na sekundę przy rozdzielczości 900×600 pikseli.

Poprawa jakości obrazu

Opracowano testowy algorytm, który konwertuje obrazy o większej głębi bitowej do 8-bitowej. Lokalnie poprawiany jest kontrast w celu skutecznej detekcji ciemnych obiektów na ciemnym tle, w tym przechodzących osób. Korzystając z oprogramowania Digital Photo Professional, każdy z obrazów zapisano w 3 nieskompresowanych, 8-bitowych wersjach:

- Z_n – zdjęcie nie przetworzone,

- Z_j – zdjęcie rozjaśnione czterokrotnie, a więc o 2 jednostki EV (ang. *exposure value*); wykorzystuje informację zawartą w dodatkowych poziomach kwantyzacji opisujących jasne fragmenty. Informacja ta w zdjęciu nieprzetworzonym jest pomijana (piksele, które miałyby mieć wartość większą niż 255 otrzymują właśnie tę wartość),
- Z_c – zdjęcie przyciemnione czterokrotnie, a więc o 2 jednostki EV; wykorzystuje informację zawartą w dodatkowych poziomach kwantyzacji opisujących ciemne fragmenty. Informacja ta w zdjęciu nieprzetworzonym jest pomijana (piksele, które miałyby mieć wartość mniejszą niż 0 otrzymują właśnie tę wartość).

Następnie każdą trójkę zdjęć reprezentującą jedną klatkę nagrania konwertowano do skali szarości uzyskując, odpowiednio, zdjęcia B_n , B_j i B_c . Dla każdego z tych zdjęć (z wyjątkiem ich pikseli brzegowych) tworzono, odpowiednio, macierze R_n , R_j i R_c , które zawierały zsumowane bezwzględne wartości różnic (liczonych w kierunku poziomym i pionowym) pomiędzy sąsiednimi pikselami tworzącymi maskę 3×3 wokół analizowanego piksela. Duże wartości w uzyskanych macierzach oznaczały dużą zmienność (a więc dużą szczegółowość) danego fragmentu zdjęcia. Miejsca prześwietlone (a więc jasne, bez szczegółów) miały więc niewielkie wartości różnic w macierzach R_j i znacznie większe w macierzach R_c (szczegóły w miejscach prześwietlonych stawały się zauważalne). Analogicznie, miejsca niedoświetlone miały niewielkie wartości w macierzach R_c i duże w macierzach R_j . Dzięki temu było możliwe określenie, które fragmenty gotowego zdjęcia mają pochodzić ze zdjęcia rozjaśnionego, a które – ze zdjęcia przyciemnionego. W tym celu tworzono macierze: D_c zawierającą dodatnie wartości różnic $R_c - R_n$ (w pozostałych przypadkach zera) oraz D_j zawierającą dodatnie wartości różnic $R_j - R_n$ (w pozostałych przypadkach zera). Obszary macierzy D_c , zawierające dodatnie wartości oznaczały, że dany fragment zdjęcia powinien być pobrany ze zdjęcia przyciemnionego, natomiast duże wartości D_j , że dany fragment powinien pochodzić ze zdjęcia rozjaśnionego. W pozostałych miejscach powinno się zastosować zdjęcie nieprzetworzone.

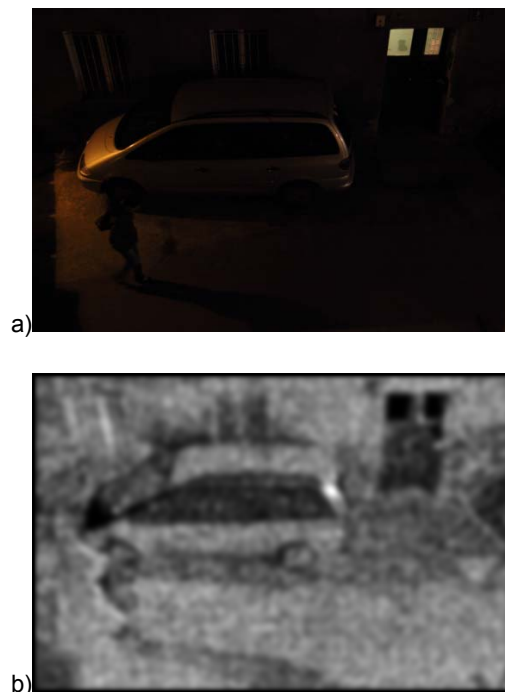
Tak utworzone zdjęcie będzie zawierać pseudokrawędzie, które wyglądają nienaturalnie i utrudniają detekcję poruszających się obiektów. Ponadto niektóre fragmenty zdjęcia mogą być sklasyfikowane jako nadające się do jednoczesnego rozjaśnienia i przyciemnienia (mają duże wartości D_j i D_c). Aby algorytm poprawnie działał wokół brzegów obiektów i nie powodował nienaturalnych rozjaśnień, należy wykryć krawędzie i ograniczyć tam jego stosowanie. W tym celu w zdjęciu przyciemnionym i rozjaśnionym stosuje się filtrację filtrem Prewitta (jednocześnie w kierunku poziomym i pionowym). Uzyskany wynik filtracji, jeśli jest większy od ustalonego progu, odejmuje się od macierzy D_j i D_c , które następnie są pierwiastkowane, aby uniknąć zbyt dużych różnic bezwzględnych pomiędzy ich sąsiednimi wartościami. Następnie są wykrywane obszary, które jednocześnie miałyby zostać rozjaśnione i przyciemnione. W tym celu macierz D_j odejmuje się od D_c a uzyskaną macierz X uśrednia filtrem uśredniającym o masce 3×3 piksele. Jej wartości, po znormalizowaniu do zakresu $\langle -1; 1 \rangle$ wskazują, jak bardzo dane miejsca mają być rozjaśnione (dla $X \in \langle 0; 1 \rangle$) lub przyciemnione (dla $X \in \langle -1; 0 \rangle$). Wartość $X = 1$

oznacza, że piksel obrazu wynikowego ma być równy wartości odpowiedniego piksela obrazu przyciemnionego, wartość $X = 0$ oznacza, że piksel ma pochodzić z obrazu nieprzetworzonego, wartość $X = -1$ oznacza, że piksel ma pochodzić z obrazu rozjaśnionego, natomiast w pozostałych przypadkach wartości pikseli są wyliczane przy wykorzystaniu liniowej zależności. Schemat algorytmu do lokalnej poprawy kontrastu zdjęcia został zobrazowany na rysunku 2.

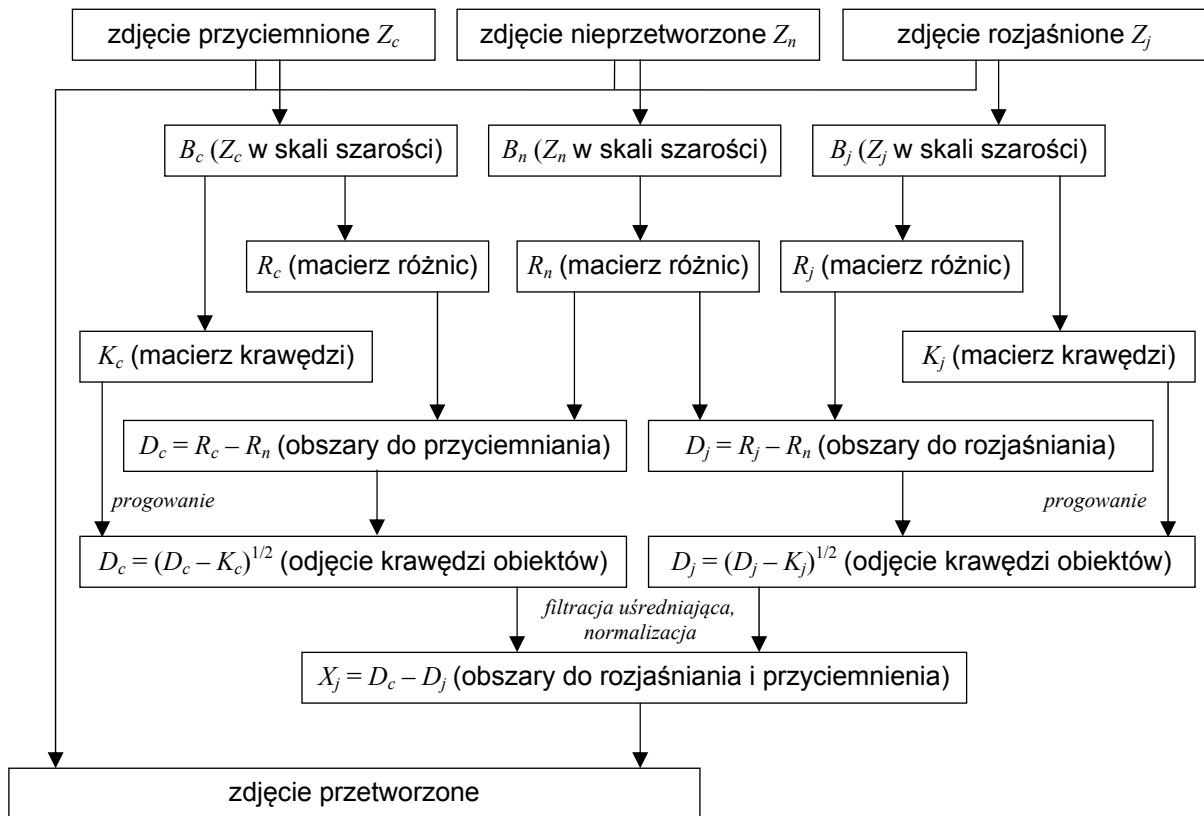
Rysunek 1 prezentuje przykładowe zdjęcie użyte w algorytmie detekcji oraz wykryte jego obszary, które powinny zostać rozjaśnione (ze względu na brak miejsc prześwietlonych, konieczność przyciemnienia nie występuje). Zaletą algorytmu jest to, że nie rozjaśnia ani nie przyciemnia obszarów, które są i tak za ciemne (np. boczne szyby pojazdu w tle) lub za jasne, by móc wykryć w nich jakiegokolwiek szczegóły. Dlatego uzyskany obraz jest bardziej realistyczny. Na rysunku 3 przedstawiono uzyskane rezultaty przetwarzania przy pomocy zaproponowanego algorytmu w porównaniu do zwiększenia jasności i zmniejszenia kontrastu zdjęcia nieprzetworzonego. Uzyskane zdjęcie zawiera więcej szczegółów, co ułatwia detekcję i śledzenie obiektu ruchomego.

Detekcja obiektów ruchomych i zliczanie osób

Algorytm detekcji obiektów ruchomych wykorzystuje model estymacji tła GMM (ang. *Gaussian mixture model*) oraz analizę BLOB. Model GMM [17] został wybrany ze względu na to, że nie odnosi się do stałego tła (jedna ramka wideo, która nie zawiera obiektów ruchomych w scenie) oraz dlatego, że jego struktura minimalizuje wpływ zmian oświetlenia w analizowanym obszarze.



Rys. 1. Od góry: zdjęcie użyte w algorytmie detekcji pieszych oraz wykryte obszary, które powinny zostać rozjaśnione przy użyciu algorytmu lokalnej poprawy kontrastu (prezentowane jasnymi odcieniami)



Rys. 2. Schemat algorytmu do lokalnej poprawy kontrastu zdjęcia

W przypadku analizy BLOB [18] parametry zostały dobrane tak, aby małe obiekty w obrazie binarnym, takie jak ptaki, psy lub szum wynikający z nagłej zmiany oświetlenia, nie były wykrywane i używane do dalszej analizy.

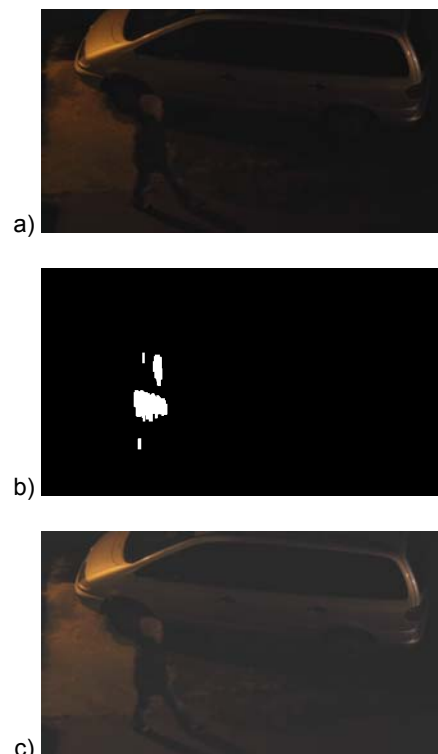
Następnym etapem algorytmu jest klasyfikacja BLOBów ze względu na ich typ (samochód, pieszy). W tym celu wzięto pod uwagę szerokość i wysokość obiektów binarnych. Początkowo autorzy pracy, w celu określenia typu obiektu, stosowali klasyfikator Bayes'a (wytrenowany przy użyciu pomiarów poszczególnych BLOBów wyrażonych w pikselach). Ostatecznie zdecydowano się na implementacyjnie prostsze i szybsze rozwiązanie – użycie współczynnika Fereta (R_f). Określa go wydłużenie cząstki (odpowiednio do uzyskanej wartości R_f – wydłużenie BLOBa w pionie lub w poziomie) i można go wyliczyć ze stosunku maksymalnej średnicy obiektu w poziomie (L_h) do maksymalnej średnicy obiektu w pionie (L_v).

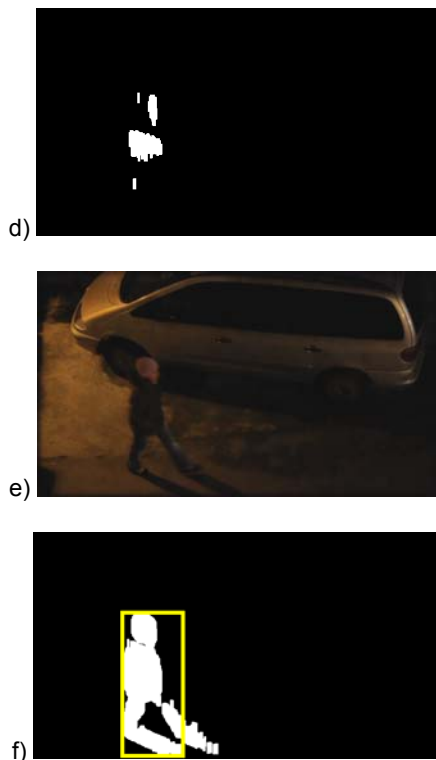
Ustalono odpowiedni próg klasyfikacji ($R_f = 1$). Obiekty o wielkości poniżej tego progu są traktowane jako piesi, a powyżej – jako pojazdy. Podczas działania programu wykrywane są wyłącznie osoby poruszające się pieszo, natomiast pojazdy nie są śledzone.

W celu śledzenia pieszych użyto filtru Kalmana [19]. Tworzy on i usuwa nowy obiekt w trakcie obecności i zniknięcia ze sceny po czasie trwania odpowiadającym 5 ramkom wideo. Model śledzenia osób poruszających się pieszo został dobrany tak, aby móc obserwować obiekty poruszające się ze stałą prędkością. Jest to opcja odpowiednia dla przewidywania pozycji osób idących pieszo w otwartej przestrzeni miejskiej.

Jak zostało przedstawione na rysunku 3, w oryginalnej sekwencji wizyjnej (przed zastosowaniem algorytmu lokalnej poprawy kontrastu) tylko część osoby została wykryta. Podobnie niską skuteczność detekcji uzyskuje się w przypadku sekwencji, która została rozjaśniona. Dobór

odpowiednich parametrów algorytmu detekcji obiektów ruchomych oraz zastosowanie metody lokalnej poprawy kontrastu umożliwiły prawidłowe wykrycia obiektu na nagraniu wykonanym przy niedostatecznym oświetleniu.





Rys. 3. Skuteczność detekcji obiektów ruchomych na pojedynczej ramce sekwencji wizyjnej: a) ramka nieprzetworzona, b) wykryty na niej obiekt ruchomy, c) ramka po rozjaśnieniu, d) wykryty na niej obiekt ruchomy, e) ramka przetworzona zaproponowanym algorytmem lokalnej poprawy kontrastu, f) wykryty na niej obiekt ruchomy

Wnioski

Autorzy zaproponowali efektywną metodę detekcji osób w przestrzeni miejskiej przy niedostatecznym oświetleniu. Algorytm lokalnej poprawy kontrastu polepsza precyzję wykrywania obiektów ruchomych przy zachowaniu realistycznych kolorów obrazu, czego nie oferują systemy wykorzystujące podczerwień. Należy zauważyć, że otrzymanie precyzyjnych wyników rozpoznawania osób wymaga doboru odpowiednich parametrów programu. W dalszej części prac autorzy przewidują zastosowanie algorytmu w systemie zliczania osób oraz implementację opracowanych rozwiązań na DSP (*digital signal processor*) [20] w celu detekcji i zliczania osób bez użycia komputera PC.

Praca wspierana ze środków DS/2015 oraz z projektu Wojewódzkiego Urzędu Pracy w Poznaniu „Wsparcie stypendialne dla doktorantów na kierunkach uznanych za strategiczne z punktu widzenia rozwoju Wielkopolski”, Poddziałanie 8.2.2 Programu Operacyjnego Kapitał Ludzki (projekt finansowany przez Unię Europejską z Europejskiego Funduszu Społecznego).

LITERATURA

- [1] Chmielewska A., Pawłowski P., Dąbrowski A., Zastosowanie metod przetwarzania sekwencji wideo do wspomagania monitoringu miejskiego, *Teoria i zastosowanie informatyki*, vol. 9, nr 3: 65-83, 2011
- [2] Balcerek J., Chmielewska A., Dąbrowski A., Jackowski D., Konieczka A., Marciniak T., Pawłowski P., Recognition of threats in urban areas by means of the analysis of video sequences, *Proc. of Multimedia Communications, Services and Security*, 41-48, Kraków, 6-7 maja 2010
- [3] Marciniak T., Dąbrowski A., Chmielewska A., Nowakowski M., Real-Time Bi-Directional People Counting Using Video Blob Analysis, *IEEE NTAV/SPA 2012, IEEE New Trends in Audio*

- and Video / Signal Processing Algorithms, Architectures, Arrangements and Applications*, 161-166, 2012
- [4] Marciniak T., Chmielewska A., Dąbrowski A., Malina A., People counting vision system based on ARM processor programmed using Simulink environment, *Electronics – constructions, technologies, applications*, No. 6, 2014, 55-59, 2014
- [5] Masoud O., Papanikolopoulos N.P., A Novel Method for Tracking and Counting Pedestrians in Real-Time Using a Single Camera, *IEEE Trans. Vehicular Technology*, vol. 50, No. 5, 1267-1278, 2001
- [6] Ye W., Zhong Z., Robust People Counting in Crowded Environment, *Proceedings of IEEE International Conference on Robotics and Biometrics*, China, 1133-1137, 2007
- [7] Kopaczewski K., Szczodak M., Czyżewski A., Krawczyk H., A method for counting people attending large public events, *Multimedia Tools and Applications*, 2013 (online). DOI: 10.1007/s11042-013-1628-0
- [8] Stahlschmidt C., Gavriilidis A., Kummert A., Density Measurements from a Top-View Position using a Time-of-Flight Camera, *International Workshop on Multidimensional Systems*, 193-198, 2013
- [9] Wang J., Kang Y., A new MSBF-based HDR image processing algorithm, *2008 IEEE Conference on Cybernetics and Intelligent Systems*, 1232-1237, 2008
- [10] Lim B.R., Park R.H., Kim S., High dynamic range for contrast enhancement, *IEEE Transactions on Consumer Electronics*, 1454-1462, 2006
- [11] Konieczka A., Balcerek J., Dąbrowski A., Metoda poprawy czytelności obrazów z monitoringu CCTV, *Elektronika – konstrukcje, technologie, zastosowania*, 77-79, nr 5/2011
- [12] Balcerek J., Konieczka A., Dąbrowski A., Stankiewicz M., Krzykowska A., Brightness Correction and Stereovision Impression Based Methods of Perceived Quality Improvement of CCTV Video Sequences, *Proc. of 4th International Conference MCSS 2011*, 64-72, 2011
- [13] Konieczka A., Balcerek J., Dąbrowski A., Iterative average filtering for image denoising, *Proc. of Signal Processing: Algorithms, Architectures, Arrangements and Applications SPA'2013*, IEEE Poland Section Chapters Signal Processing, Circuits and Systems, 302-305, Poznań, Poland, 26-28 September 2013
- [14] Konieczka A., Balcerek J., Dąbrowski A., Iteracyjna metoda usuwania szumu impulsowego z obrazów cyfrowych, *Elektronika – konstrukcje, technologie, zastosowania*, 21-23, nr 4/2014
- [15] Konieczka A., Balcerek J., Dąbrowski A., Improvement of dynamic range of photography in RAW format representation, *Proc. of Signal Processing SPA'2009*, IEEE Poland Section Chapters Signal Processing, Circuits and Systems, 62-65, Poznań, Poland, 24-26 September 2009
- [16] Konieczka A., Balcerek J., Dąbrowski A., Powiększanie zakresu dynamicznego zdjęć zapisanych w formacie RAW, *Elektronika – konstrukcje, technologie, zastosowania*, 42-45, nr 3/2010
- [17] Zivkovic Z., Improved adaptive Gaussian mixture model for background subtraction, *Proceedings of the 17th International Conference on Pattern Recognition*, 28-31, 2004
- [18] Chen T.H., Lin Y.F., Chen T.Y., Intelligent Vehicle Counting Method Based on BLOB Analysis in Traffic Surveillance, *Second International Conference on Innovative Computing*, 238-242, 2007
- [19] Huang S., Hong J., Moving object tracking system based on camshift and Kalman filter, *International Conference on Consumer Electronics, Communications and Networks*, 1423-1426, 2011
- [20] Chmielewska A., Weychan R., Marciniak T., Dąbrowski A., Hartwich M., Owczarczak M., Fast Prototyping for Video Monitoring Systems with the Use of DSP Module, *International Journal of Electronics and Telecommunications*, Volume 59, Issue 4: 375-381, 2013

Autorzy: mgr inż. Adam Konieczka, Politechnika Poznańska, Wydział Informatyki, ul. Piotrowo 3a, 60-965 Poznań; mgr inż. Agata Chmielewska, Politechnika Poznańska, Wydział Informatyki, ul. Piotrowo 3a, 60-965 Poznań, E-mail: chmielewska.agata@gmail.com; mgr inż. Julian Balcerek, Politechnika Poznańska, Wydział Informatyki, ul. Piotrowo 3a, 60-965 Poznań.