



Krzysztof Kowalak, Sławomir Maćkowiak

# TECHNIKI ZLICZANIA OSÓB

## WYKORZYSTUJĄCE OBRAZ JAKO INFORMACJĘ WEJŚCIOWĄ

Zliczanie osób w systemach analizy obrazu ma wiele zastosowań. Może zapewnić wiele korzyści biznesowych i operacyjnych różnym organizacjom, zarówno w zakresie optymalizacji sposobu działalności, jak i w sposobie pozyskiwania danych statystycznych, np. dotyczących frekwencji osób.

Zapewne większość zna doniesienia dotyczące różnic w podawaniu frekwencji uczestników zbiorowych marszy. Różnice te mogą wynikać z przyjętych odmiennych metodologii określania liczebności tłumu, ale też z zastosowanych narzędzi technicznych. Często w dyskusjach jest podnoszona możliwość policzenia osób i oszacowania tłumu z wykorzystaniem analizy obrazu. Trzeba mieć jednak na względzie, że także w systemach analizy obrazu stosuje się różne techniki zliczania, z licznymi ograniczeniami, przeznaczone do odmiennych zastosowań.

Wybór techniki zliczania zależy od zastosowania, ponieważ ta sama technika w przypadku analizy pojedynczych osób charakteryzuje się inną dokładnością zliczania, niż w przypadku

analizy tłumu. Podobna sytuacja z różnicami w osiągniętych dokładnościach występuje również przy różnym sposobie montażu kamer, np. kamera podwieszona z obiektywem skierowanym ku dołowi lub kamera ulokowana pod kątem w stosunku do obserwowanej sceny.

Analizatory wizyjne w powszechnych rozwiązaniach umożliwiają zastosowanie tzw. wirtualnych linii zliczania (rys. 1a i rys. 2a). Pełnią one funkcję liczników dwukierunkowych, analizując ruch obiektów widzianych na obrazie. Przecięcie umownej linii podczas ruchu obiektu o ściśle zdefiniowanym rozmiarze w określonym kierunku powoduje zwiększenie stanu licznika. Takie analizatory często pozwalają zdefiniować jednocześnie wiele umownych linii na jednym kanale wizyjnym, rozpoznając kierunek ruchu obiektu przecinającego którąś z tych linii. Drugą możliwością jest zdefiniowanie na obrazie wirtualnych zamkniętych stref (rys. 1b i rys. 2b), w których dokonuje się zliczania osób w danej chwili.

Zastosowanie wirtualnych linii sprawdza się przy określaniu liczby osób wchodzących

w obszar ograniczony fizycznie i wychodzących z niego, np. zamknięte obiekty, sklepy, miejsca użyteczności publicznej, strefy, w których można zdefiniować miejsca wejść i wyjść. Zastosowanie wirtualnych stref natomiast pozwala uzyskać dane z obszaru obserwowanego przez kamerę bez potrzeby określania miejsc wejść i wyjść.

Wirtualne linie zliczania wymagają określenia parametru reprezentującego obiekt, i śledzenia toru jego poruszania się w scenie. Może nim być np. czubek głowy lub podstawa nóg, w zależności od zastosowanego algorytmu i sposobu montażu kamery. W przypadku wirtualnych stref często dokonuje się estymacji gęstości tłumu i szacowania jego rozmiaru na podstawie zdefiniowanych cech charakterystycznych, tym samym obiektu nie analizuje się pojedynczo, a raczej masowo, w przeciwieństwie do metod stosowanych w przypadku wirtualnych linii.

Łatwiejszą analizą, z punktu widzenia dokładności szacowania wyników, są metody dedykowane pojedynczym osobom lub tłumowi

# POZNAJ INTELIGENTNĄ ANALIZĘ OBRAZU



IWA znacznie rozszerza możliwości znanej i stosowanej od lat detekcji ruchu, pozwalając na lepszą i automatyczną obsługę krytycznych zdarzeń.

#### Wybrane funkcje IVA:

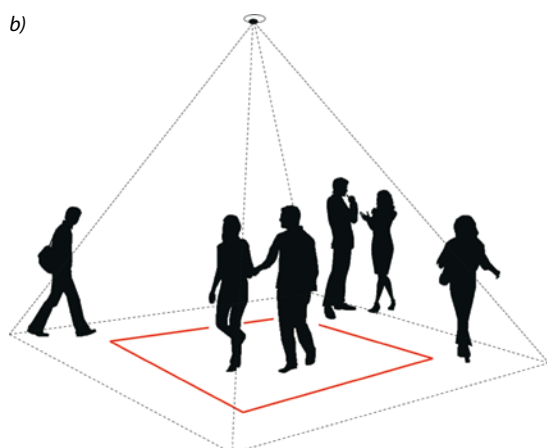
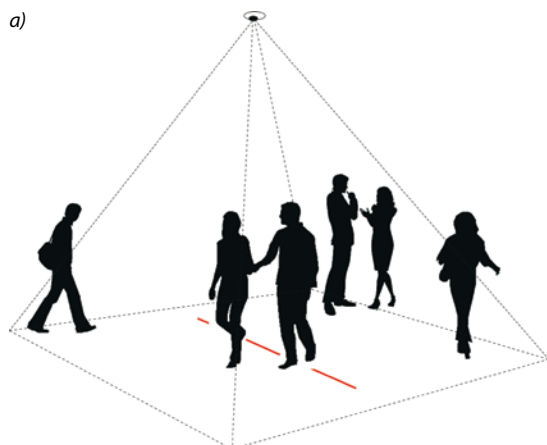
- diagnostyka wideo,
- wirtualna granica,
- zastrzeżony obszar,
- pozostawienie lub zniknięcie obiektu,
- wykrywanie nietypowego zachowania,
- nieprawidłowe parkowanie,
- detekcja tłumu,
- wykrywanie biegnięcia.

Więcej informacji znajdziesz na [www.tiandy.pl](http://www.tiandy.pl)

Kamery z funkcją IVA  
już od **199zł**

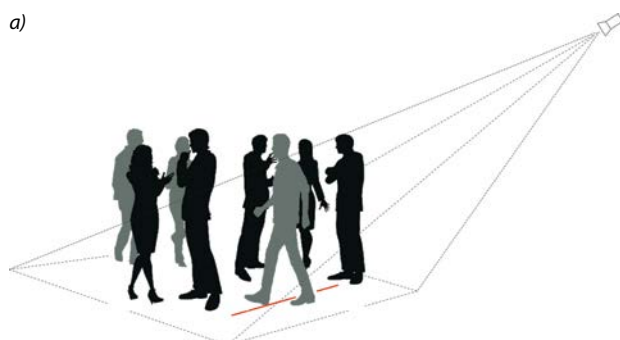


Genway  
Oficjalny dystrybutor marki Tiandy  
Tel.: (24) 264 77 33  
Email: [info@genway.pl](mailto:info@genway.pl)  
[www.genway.pl](http://www.genway.pl)



Rys. 1. Kamera skierowana prostopadle w dół:

a) określona wirtualna linia (kolor czerwony), której przekroczenie powoduje zwiększenie licznika,  
b) określona wirtualna strefa (kolor czerwony), w której następuje zliczanie osób



Rys. 2. Kamera skierowana pod kątem:

a) określona wirtualna linia (kolor czerwony), której przekroczenie powoduje zwiększenie licznika,  
b) określona wirtualna strefa (kolor czerwony), w której następuje zliczanie osób.

W obu przypadkach występuje przesłanianie się obiektów, co wpływa na detekcję osób na podstawie fragmentów sylwetek

o niewielkiej gęstości. Sceny zatłoczone są wyzwaniem w widzeniu maszynowym ze względu na przesłonięcia między obiektami wynikające z dużej gęstości tłumu, niedostateczne możliwości obserwacji wszystkich części ciała, małą rozdzielność cech w wyglądzie głowy, podobieństwo głowy np. do fragmentów ramion osoby będącej w pobliżu, duża różnorod-

ność w ubiorze osób. W przypadku obserwacji tłumu zwykle spotykamy się z kadrami o szerokim ujęciu, w których postacie są stosunkowo małe, zmiennymi warunkami oświetlenia i brakiem możliwości wydzielenia obiektów z tła. Analiz dokonuje się ławiej, gdy osoby się poruszają, gdy można je wyróżnić na obrazie tła lub gdy sylwetka osób jest duża w stosunku do rozmiarów

obrazu, a rozdzielczość przestrzenna obrazów pozwala na rozróżnianie cech charakterystycznych osób.

Także w przypadku analizy pojedynczych osób pojawiają się problemy, które mogą fałszować wynik, np. cienie osób mogą zostać potraktowane jako kolejna osoba. W takiej sytuacji należy inaczej skierować oświetlenie lub dodatkowo doświetlić przejścia z innych kierunków, żeby obserwowane obiekty były dobrze widoczne. Interesującymi rozwiązaniami są także dedykowane algorytmy usuwania cienia.

W literaturze oraz w badaniach naukowych przyjęto podział metod zliczania osób na dwie kategorie technik, które odzwierciedlają ich potencjalne możliwości zastosowań, choć nie ograniczają ich użycia tylko do nich. Metody podzielono ze względu na możliwości detekcji i analizy pojedynczych osób lub szacowania rozmiaru tłumu. Wydzielono metody detekcji oparte na modelu sylwetki ciała ludzkiego i metody oparte na mapach odwzorowań. Te pierwsze wykorzystują segmentację obrazu i detekcję charakterystycznych fragmentów ciała: głowy, ramię, torsu czy pozostałych kończyn. Drugie polegają na powiązaniu osób z cechami lokalnymi lub globalnymi pozyskiwanymi z obrazu, dzięki czemu są w stanie estymować liczbę osób znajdujących się na obrazie bez detekcji i klasyfikacji sylwetek ciał.

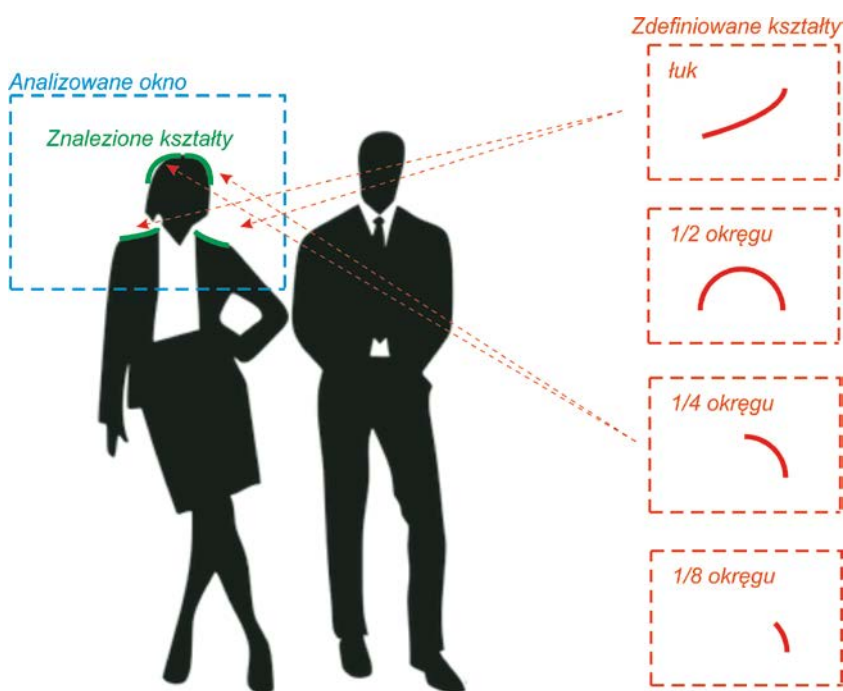
#### METODY DETEKЦИИ OSÓB OPARTE NA MODELU SYLWETKI CIAŁA

Jednym z ciekawszych rozwiązań dotyczących zliczania osób w obrazie jest grupa metod detekcji osób oparta na kształcie wydzielanej w obrazie sylwetki ciała. Kształt może się odnosić do geometrii konturu obiektu (w dwóch wymiarach) lub fragmentu konturu reprezentującego fragment kończyny.

W systemach dozoru wizyjnego często obiekty przesłaniają się wzajemnie lub są zasłonięte przez samochody bądź elementy drobnej architektury. W takich sytuacjach detekcji można dokonać wyłącznie na podstawie widocznych fragmentów sylwetki osób.

Ta grupa metod detekcji sprawdza się w przypadku typowego montażu kamer dozoru wizyjnego obserwujących scenę pod kątem, jednak ich skuteczność mocno zależy od rozmiarów obserwowanych obiektów i stopnia ich wzajemnego przesłaniania się. Radzą sobie lepiej, gdy rozmiar rejestrowanych obiektów jest znaczący w stosunku do wymiarów rejestrowanego obrazu. Niestety ich wadą jest duża złożoność wymagająca pokaznych obliczeń. W tych metodach dokonuje się analizy znalezionych części ciała i próby dopasowania kilku różnych części do jednej postaci na podstawie obliczonego prawdopodobieństwa połączeń między wydzielonymi fragmentami sylwetki ciała.

Dobry opis kształtu ułatwia prowadzenie rozpoznania sylwetek, a co za tym idzie istnieje możliwość, na podstawie podziału sylwetki na naturalne części, dokonania reprezentacji sylwetki za pomocą zbioru ele-



Rys. 3. Sposób wyznaczania kształtów

mentów reprezentujących kończyny podmiotu. Przyjrzyjmy się zatem dokładniej tej grupie rozwiązań.

System dokonuje detekcji osób w obrazie, analizując wyniki klasyfikacji poszczególnych części ciała i prawdopodobieństwo ich połączenia ze sobą. Rozpoznawane części ciała przez algorytm to głowa z górną częścią tułowia, tors i nogi. Na podstawie obserwacji konturów ludzkiego ciała opracowano specjalną grupę obiektów reprezentującą możliwe lokalne kształty, za pomocą których można modelować sylwetkę osoby.

Wyróżnia się cztery podstawowe kształty: linia, 1/2 okręgu, 1/4 okręgu, 1/8 okręgu oraz ich symetryczne pary (rys. 3). Każdy kształt może być umieszczony na obrazie pod różnym kątem. Pojedynczy obiekt (kształt) zawiera się pomiędzy 4. a 12. punktami obrazu, zatem na obszarze o rozmiarach 24 x 58 punktów algorytm musi przeanalizować ponad 850 tys. kształtów.

Na podstawie obrazu wejściowego jest tworzony obraz z zarysem krawędzi. Są to dane wejściowe podawane na detektory poszczególnych, zdefiniowanych kształtów i następnie poddawane klasyfikacji.

Uczenie systemu odbywa się z wykorzystaniem algorytmu wzmacniania (*boosting*). Jego idea polega na budowaniu mocnego i złożonego klasyfikatora ze słabych i prostych klasyfikatorów. Słabe algorytmy uczące, których skuteczność jest niewiele lepsza niż losowe zgadywanie, mogą być wykorzystane do utworzenia skutecznego, silnego klasyfikatora.

Dla każdego kształtu jest przygotowany jeden klasyfikator. Następnie tworzy się jeden silny klasyfikator z puli dostępnych słabych klasyfikatorów. Silny klasyfikator to liniowa

kombinacja klasyfikatorów słabych. Ustala się próg oddzielający dobrze sklasyfikowane obiekty od sklasyfikowanych niepoprawnie w taki sposób, aby zawierał jak najwięcej pozytywnie sklasyfikowanych próbek i odrzucał maksymalnie dużo niepoprawnie sklasyfikowanych. Próbkę pozytywnie sklasyfikowaną służą do nauki silnego klasyfikatora kolejnej warstwy w algorytmie AdaBoost. AdaBoost, czyli tzw. adaptacyjny *boosting*, polega na wywoływaniu wybranego słabego algorytmu uczącego w serii kilku iteracji i skierowaniu uwagi słabego klasyfikatora na pewne elementy (trudne do wyuczenia) ze zbioru treningowego.

Wróćmy do procesu detekcji. Badana próbka trafia na blok detektorów. W wyniku ich działania można oczekiwać wyboru jednej z następujących możliwości: tylko jeden z detektorów pozytywnie rozpoznał obiekt (zdefiniowany kształt), kilka detektorów równoległe rozpoznało obiekt lub żaden z detektorów nie rozpoznał obiektu. Wynikiem działania detektorów są znalezione poszczególne fragmenty ciała oraz, jeśli to możliwe, całe ciała (rys. 4). W procesie detekcji każda próbka jest także poddawana skalowaniu, aby rozróżnić obiekty o różnych wymiarach.

Aby odpowiednio połączyć części ciała należące do tej samej osoby, wykorzystuje się klasyfikator oparty na prawdopodobieństwie, że dany obiekt  $x$  należy do klasy  $k$ . Dla danego obiektu  $x$  oblicza się prawdopodobieństwo przynależności kolejno do każdej z klas, następnie wybiera klasę najbardziej prawdopodobną.

W praktycznej realizacji na podstawie znalezionych fragmentów ciała zakłada się hipotetyczną liczbę osób na obrazie. Odpowiada ona liczbie zidentyfikowanych fragmentów



Rys. 4. Wynik detekcji części ciała  
(kolor żółty – całe ciało, kolor czerwony – ramiona  
z głową, kolor fioletowy – tors, kolor niebieski – nogi)

będących głowami i całymi ciałami. Jedna hipotetyczna osoba to jedna zidentyfikowana głowa lub jedna zidentyfikowana sylwetka. W następnym kroku tworzy się macierz zawierającą wszystkie możliwe części ciała, jakie mogą być zidentyfikowane, biorąc pod uwagę liczbę hipotetycznych osób (części te stanowią kolumny macierzy) oraz wszystkie części ciała sklasyfikowane w poprzednim etapie (części te stanowią wiersze macierzy). Macierz zawiera wartości reprezentujące prawdopodobieństwa połączeń pomiędzy położeniem hipotetycznych części ciała ze znanymi częściami ciała. Kryterium połączenia może być dopasowanie kształtów lub innych cech obrazowych.

W kroku kolejnym skreśla się kolejno kolumnę i rząd, pomiędzy którymi prawdopodo-

bieństwo jest małe. Po tej operacji otrzymuje się trzy grupy: znalezione części ciała dopasowane do hipotetycznych części ciała, znalezione części ciała niemające odpowiedników w hipotetycznych częściach ciała oraz hipotetyczne części ciała, które nie zostały dopasowane do znalezionych części ciała. Następnie sortuje się hipotetyczne osoby względem odległości od kamery.

Ponieważ osoby znajdujące się najbliżej kamery nie są przysłonięte przez inne osoby, algorytm powinien zidentyfikować wszystkie części ich ciała. Wraz ze wzrostem odległości od kamery osoby są bardziej przysłonięte, więc mogą zostać zidentyfikowane tylko ich głowy. Analizując hipotetyczne osoby w kolejności rosnącej odległości od kamery (im większy obiekt, tym potencjalnie bliżej znajduje się on kamery), za pomocą klasyfikatora opartego na prawdopodobieństwie połączeń określa się, czy dana znaleziona część ciała należy do danej hipotetycznej osoby. Po złożeniu znalezionych fragmentów ciała dopasowanych do odpowiednich osób otrzymuje się liczbę osób w obrazie.

Niestety metoda ta nie jest wolna od wad. Bardzo często fragmenty tła są klasyfikowane jako kształty ciała, co skutkuje fałszywymi detekcjami, a tym samym wynik zliczania nie jest poprawny.

Do nauki algorytmu wymaga się po kilka tysięcy obrazów z postaciami ustawionymi

wprost do kamery oraz kilka tysięcy obrazów przedstawiających osoby ustawione bokiem. Do każdego poziomu uczenia algorytmu AdaBoost wykorzystuje się po kilka tysięcy „błędnych próbek”, np. przedstawiających tło. Mimo to wyznaczanie osób na podstawie zarysów krawędzi jest dość trudne, szczególnie w przypadku sekwencji zatłoczonych lub wielu elementów tła. Wtedy często zarysy danych części ciała mogą powstać przez przypadkowo ustawione osoby bądź przedmioty.

Alternatywną metodą rozpoznawania sylwetek osób na podstawie kształtu jest wyznaczanie konturu sylwetki i jego podziału na naturalne części. W pierwszym kroku dokonuje się wydzielenia obiektów ruchomych z tła, następnie dla obszaru wydzielonego wyznacza się otoczkę wypukłą rozpiętą na obszarze wydzielonym oraz punkty wklęsłości w konturze. W kolejnym kroku z wyznaczonych punktów wklęsłości prowadzi się cięcia służące do podzielenia sylwetki na fragmenty naturalne. Na ich podstawie będzie prowadzona dalsza klasyfikacja sylwetek.

Mając obszary podzielone na fragmenty, można z tych fragmentów sylwetki wygenerować wiele scenariuszy (hipotez) ich złożenia w podstawowy model ciała. Struktura takiego modelu zakłada, że sylwetka osoby składa się z torsu oraz przyłączonych do niego głowy, rąk i nóg.

**PRS - bezpłatny dodatek do rozpoznawania tablic rejestracyjnych**  
minimalne wymagania dla PRS ALNET - NetStation 8 lub wyższy

Ponad 200 000  
systemów na świecie  
najnowsze referencje:



Sieć sklepów Auchan Rosja  
2500 kanałów IP



Państwowe Koleje Łotewskie  
6500 kanałów IP



Komisja Europejska Luksemburg  
1300 kanałów IP

[www.alnetsystems.com](http://www.alnetsystems.com) [www.youtube.com/alnetsystems](http://www.youtube.com/alnetsystems)

Przed utworzeniem listy hipotez należy wyznaczyć wszystkie możliwe kombinacje takiego modelu. Proces tworzenia różnych scenariuszy zakłada, że każdej części modelu ciała można przypisać każdy wykryty fragment konturu. Możliwe są również scenariusze, że dana część modelu ciała nie ma przypisanego konturu, lub odwrotnie, wykryty kontur nie jest skojarzony z żadną częścią modelu. Liczba różnych kombinacji złożenia modelu może być więc bardzo duża. W celu jej ograniczenia można wykorzystać fakt, że nie wszystkie wykryte kontury nadają się do reprezentacji poszczególnych fragmentów modelu ciała. Główną cechą decydującą o dyskryminacji danego fragmentu konturu może być stosunek wysokości do szerokości.

Znając listę wszystkich możliwych kombinacji podstawowego modelu ciała, można przystąpić do obliczania wektorów hipotez. Zawierają one informację, czy dana część modelu ma przyporządkowany fragment konturu i czy widok modelu jest z przodu (tyłu), czy z boku. Na podstawie wektora hipotez oraz pewnych cech opisujących części modelu ciała można wyznaczyć miarę pozwalającą na zaklasyfikowanie danego obiektu jako osoby. Spośród cech charakteryzujących poszczególne części modelu ciała należy wyróżnić: stosunek szerokości do długości (proporcje), współrzędne położenia fragmentu względem części odniesienia (torsu), stosunek długości danej części do długości innej części.

Parametry rozkładów przytoczonych cech pozwalają określić, czy wykryty i zamodelowany obiekt może być osobą.

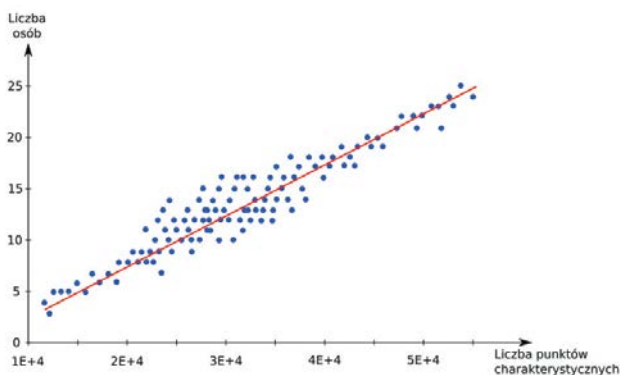
W systemach dozorowych, w których zastosowano kamery głębi bądź kamery stereoskopowe, gdzie głębia jest wyznaczana, metody oparte na detekcji charakterystycznych części ciała są szeroko wykorzystywane. Istotną ich cechą jest zwiększenie dokładności zliczania osób dzięki zminimalizowaniu wpływu przesłaniania się obiektów na podejmowane decyzje.

W przypadku kamer głębi zamontowanych pod niewielkim kątem do kierunku poruszania się osób często stosuje się wsparcie informacji o charakterystycznym kształcie, np. ramion i głowy, informacją z mapy głębi mówiącą o wypukłości sylwetki. Pozwala ona precyzyjniej określić liczbę osób w obrazie. Opisany system obejmuje trzy podstawowe kroki. Pierwszym jest usuwanie tła, drugim – detekcja charakterystycznych kształtów ciała, np. poprzez dopasowanie odpowiednich punktów na podstawie szablonu dopasowania (tors, ramiona, głowa). Można tutaj zastosować wcześniej opisaną metodę detekcji i klasyfikacji kształtów.

Kolejnym krokiem jest określenie dokładnej otoczki na podstawie informacji o wypukłości punktów w mapie głębi i wybranie takich, które zawierają kształty charakterystyczne ze zdefiniowanego wcześniej zbioru oraz wskazanie pojedynczego punktu w tej otoczce, który znajduje się najwyżej. Jest nim czubek głowy lub nakrycie głowy, jeśli przypomina



Rys. 5. Obrazy z bazy danych PETS2009, z zaznaczonymi przykładowymi punktami charakterystycznymi. Kolorem czerwonym oznaczono punkty charakterystyczne należące do elementów tła, zielonym punkty charakterystyczne należące od obiektów przemieszczających się w obrazie



Rys. 6. Wykres obrazujący zależność między liczbą osób a liczbą punktów charakterystycznych wykrytych w obrazie. Niebieskie punkty oznaczają rzeczywistą liczbę osób w scenie w zależności od liczby punktów, natomiast czerwona linia jest zależnością przybliżoną

ono charakterystyczny zdefiniowany kształt, i prawdopodobieństwo połączenia z innymi fragmentami ciała było wysokie. Tak określony punkt jest punktem śledzonym w obszarze i informacją, na podstawie której dokonuje się zliczenia osób.

W przypadku kamer głębi lub kamer stereoskopowych montowanych na suficie z obiektywem skierowanym pionowo w dół stosuje się algorytmy segmentacji mapy głębi. Mapa głębi reprezentuje odległości od kamery do każdego obiektu przemieszczającego się pod kamerą. Osoby o różnym wzroście, zwłaszcza równoczesne występowanie wielu osób we fragmencie obserwowanej sceny, powodują powstawanie na mapie głębi lokalnych zmian odległości między kamerą a obiektami. Obraz mapy głębi jest traktowany jako powierzchnia topograficzna. Mapy głębi wskazują lokalne ekstrema. Najmniejsze wartości odległości wskazują czubek głowy. W prosty sposób interpretując takie mapy, można określić liczebność w danym obszarze.

#### METODY DETEKcji OSÓB OPARTE NA MAPACH ODWZOROWAŃ

Drugą grupę metod służących do zliczania ludzi w obrazie stanowią metody oparte na mapach odwzorowań. Odwzorowują one powiązania między ludźmi a lokalnymi lub globalnymi cechami charakterystycznymi pozytywnymi z obrazu, dzięki czemu są w stanie oszacować liczbę osób znajdujących się na obrazie. W przeciwieństwie do metod opartych na modelu sylwetki ciała nie wymagają detek-

cji ani poszczególnych sylwetek, ani ich części. Cechy charakterystyczne wykorzystywane w tego rodzaju metodach bazują na gradientach (SIFT, SURF, punkty narożników), punktach poruszających się w obrazie, bądź cechach teksturalnych. Mogą być wyznaczone zarówno dla całego obrazu, jak i pewnego obszaru zainteresowania wyznaczonego przez wirtualną strefę.

W metodach opartych na mapach odwzorowań często wykorzystuje się istnienie w przybliżeniu liniowej zależności między liczbą punktów charakterystycznych reprezentujących obiekty w obrazie a liczbą osób, które się na tym obrazie znajdują (rys. 6). Na tej podstawie można oszacować liczbę osób, posługując się prostym równaniem:

$$N_p = N_c / N_l \quad (1)$$

gdzie:  $N_p$  jest liczbą ludzi widocznych w obrazie,  $N_c$  liczbą punktów charakterystycznych wyznaczonych w obrazie, natomiast  $N_l$  liczbą punktów charakterystycznych przypadających na jedną osobę.

Ostatni z opisywanych parametrów musi zostać podany wcześniej. Do jego wyznaczenia najlepiej wykorzystać informacje o liczbie punktów charakterystycznych i rzeczywistej liczbie osób w obrazie o średnim lub dużym natężeniu tłumu. Równanie (1) może wydawać się trywialne, jednak jest ono dosyć dokładne, uwzględniając małe nakłady obliczeniowe wymagane do uzyskania wyniku.

Oprócz wysokiej dokładności zliczania w większości przypadków jest również istotne działanie algorytmu zliczania w czasie rzeczywistym. Dlatego stosowane rozwiązania muszą być na tyle mało złożone, by nie dochodziło do pomijania kolejnych ramek obrazu rejestrowanych przez kamerę. Jednym z ciekawszych rozwiązań zapewniających wysoką skuteczność zliczania przy stosunkowo małym nakładzie obliczeniowym jest metoda wykorzystująca cechy charakterystyczne w formie narożników wykrytych w obrazie i wektorów ruchu dla tych punktów. Detekcja narożników odbywa się na podstawie dużej wartości gradientu w obrazie i dużej krzywizny krawędzi.

Informacje o narożnikach są bardziej użyteczne od informacji pozyskanych np. na gładkich obszarach lub prostych krawędziach. Wykryte narożniki są punktami charakterystycznymi, stanowiącymi podstawę prezentowanego rozwiązania. Dla każdego punktu charakterystycznego wyznacza się wektor ruchu za pomocą techniki pasowania bloków o różnych rozmiarach. Punkty charakterystyczne, dla których wektory ruchu są niezerowe, czyli takie, które zmieniają swoje położenie w obrazie, są traktowane jako punkty zainteresowania, które należą do obiektów przemieszczających się w scenie. Pozostałe są traktowane jako punkty należące do elementów składowych tła. Do oszacowania liczby osób w scenie wykorzystuje się równanie (1).

Metody bazujące na mapach odwzorowań charakteryzują się łatwością implementacji

oraz działaniem w czasie rzeczywistym. Odbywa się to kosztem dokładności zliczania, która może czasem wynosić od kilku do kilkunastu osób, w zależności od natężenia ruchu w scenie. W celu odseparowania punktów charakterystycznych obiektów zainteresowania od punktów stanowiących elementy tła konieczne jest modelowanie tła, a co za tym idzie, niezbędna jest informacja o ruchu w scenie. W rezultacie może nie dochodzić do detekcji osób, które znajdują się w scenie, ale się nie przemieszczają, przez co są traktowane jako elementy tła.

Zliczanie ludzi w tłumie również stanowi nie lada wyzwanie. Im więcej osób znajduje się w scenie, tym częściej dochodzi do przesłonięć, które zmieniają liczbę punktów charakterystycznych wykrywanych na obrazie. W przypadku obserwacji dużej sceny trzeba się również liczyć z wpływem perspektywy na działanie algorytmów wyznaczania punktów charakterystycznych. Osoby znajdujące się daleko od kamery są mniejsze od osób, które znajdują się blisko obiektywu, a wraz z rozmiarem rejestrowanych obiektów może zmieniać się liczba wykrytych punktów (im mniejszy obiekt, tym mniej punktów charakterystycznych może zostać wykrytych). Błąd wynikający z perspektywy można znacznie zmniejszyć, dzięki uwzględnieniu współczynnika zmiany liczby punktów charakterystycznych w zależności od ich położenia w scenie.

Ze względu na to, że metody należące do opisywanej grupy bazują na szacowaniu liczby osób na podstawie liczby punktów charakterystycznych w obrazie, nie ma informacji

o lokalizacji poszczególnych osób w scenie. By to było możliwe, należy przeprowadzić grupowanie punktów charakterystycznych. Wykorzystuje się w tym celu algorytm iteracyjny, który rozpoczynając pracę z dwa razy większą liczbą potencjalnych grup niż oszacowana liczba osób w obrazie, przyporządkowuje prawdopodobieństwo wiążące punkty charakterystyczne z każdą grupą. Następnie jest przeprowadzana maksymalizacja podobieństwa w odniesieniu do parametrów modelu grupy.

Parametry modelu grupy są aktualizowane, łącznie z lokalizacją grupy i prawdopodobieństwem jej wystąpienia. Pozwala to na określenie położenia poszczególnych sylwetek osób na podstawie miejsca punktów charakterystycznych w obrazie.

Trudno jednoznacznie odpowiedzieć na pytanie, która metoda zliczania ludzi w obrazie jest najlepsza. Zależy to od wielu czynników, takich jak obserwowany obszar, zastosowanie, dokładność zliczania. Żadna nie jest metodą uniwersalną, którą można by zastosować w każdej sytuacji. Konieczne jest więc odpowiednie dobranie metody do wymagań i warunków, w jakich przewiduje się jej działanie.

**Sławomir Maćkowiak** oraz **Krzysztof Kowalak** są pracownikami naukowymi Katedry Telekomunikacji Multimedialnej i Mikroelektroniki Politechniki Poznańskiej kierowanej przez prof. Marka Domańskiego. W swojej pracy badawczej zajmują się zagadnieniami zastosowania zaawansowanej analizy obrazów w inteligentnych systemach dozoru wizyjnego. ●

## VCA – PRZEGLĄD OFERTY RYNKOWEJ

### VCA

#### Aimetis Symphony Analytics



#### SUMA solutions sp. z o.o.

biuro@suma.com.pl; www.suma.com.pl

Aimetis Corp. od wielu już lat oferuje klientom szeroki wybór funkcji analizujących obraz wideo i dostarczających cennych informacji wspierających bezpieczeństwo obiektów i ludzi. Aimetis Analytics to idealne rozwiązanie dla przedsiębiorstw, którym zależy na bardziej efektywnym i inteligentnym systemie monitoringu wizyjnego. W najnowszej odsłonie oprogramowania do zarządzania systemami dozoru wizyjnego Symphony 7 został wprowadzony nowy model licencjonowania VCA. Inteligentna analiza treści obrazu jest dostępna w formie pojedynczych pakietów, które można dokupić do każdego rodzaju licencji Symphony 7.

Aimetis Analytics obejmuje takie typy zdarzeń, jak wykrywanie oraz śledzenie ludzi i pojazdów w otwartej przestrzeni, wykrywanie pozostawionego lub usu-

niętego przedmiotu, rysowanie map ciepła, wykrywanie tzw. włączegostwa osób pozostających w pomieszczeniach. Oprogramowanie umożliwia wykrywanie osób wkraczających na określony teren. Ponadto Aimetis oferuje oprogramowanie do rozpoznawania tablic rejestracyjnych (LPR).

Do znanych i przez lata sprawdzonych rozwiązań Aimetis dodał dwie interesujące funkcjonalności: rozpoznawanie twarzy oraz wykrywanie grup ludzi. Umożliwiają one monitorowanie zajętości wybranej przestrzeni przez gromadzących się ludzi. Dzięki Aimetis Analytics klasyczny i pasywny system monitoringu wizyjnego przekształca się w nowe narzędzie o większych możliwościach. Otrzymane z oprogramowania dane mogą być wykorzystane także przez działy sprzedaży i marketingu.