(19) URZĄD PATENTOWY RZECZYPOSPOLITEJ POLSKIEJ

(10) **PL/EP 3534294 T3**

(12) # Tłumaczenie patentu europejskiego

(96) Data i numer zgłoszenia patentu europejskiego:
2018.03.02 18159762.6

(97) O udzieleniu patentu europejskiego ogłoszono:
2023.09.06 Europejski Biuletyn Patentowy 2023/36
EP 3534294 B1

(43) Data publikacji o zgłoszeniu:
2019.09.04 Europejski Biuletyn Patentowy 2019/36

(45) Data publikacji o złożeniu tłumaczenia patentu:
2024.01.15 Wiadomości Urzędu Patentowego 2024/03

(51) MKP:

G06V 20/52	*(2022.01)*
G06V 10/82	*(2022.01)*
G06V 10/44	*(2022.01)*
G06V 10/764	*(2022.01)*

(73) Uprawniony:
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Kraków, PL

(72) Twórca(-y):
MICHAŁ GREGA, Kraków, PL
ANDRZEJ DZIECH, Kraków, PL
KRZYSZTOF RUSEK, Lipnica Wielka, PL
ANDRZEJ MATIOLAŃSKI, Kraków, PL
PIOTR GUZIK, Krosno, PL

(74) Pełnomocnik:
rzecz. pat. Marek Bury, Warszawa, PL

(54) Tytuł:
Realizowany komputerowo sposób przetwarzania sekwencji klatek, urządzenie do cyfrowego przetwarzania klatek, produkt w postaci programu komputerowego i system monitorowania

PL/EP 3534294 T3

Realizowany komputerowo sposób przetwarzania sekwencji klatek, urządzenie do cyfrowego przetwarzania klatek, produkt w postaci programu komputerowego i system monitorowania

Opis

[0001] Wynalazek dotyczy realizowanego komputerowo sposobu przetwarzania sekwencji klatek w celu wykrywania niebezpiecznych obiektów, urządzenia do cyfrowego przetwarzania klatek, produktu w postaci programu komputerowego oraz systemu monitorowania z automatycznym ulepszeniem wykrywania zdarzeń związanych z użyciem niebezpiecznych obiektów. Przez niebezpieczny obiekt można rozumieć pistolet, siekierę, nóż, karabin lub inne potencjalnie niebezpieczne narzędzie, które nie jest normalnie używane na monitorowanym obszarze.

[0002] W dokumencie patentowym USA nr US6028626 ujawniono system nadzoru posiadający co najmniej jedną główną kamerę wideo do przekształcania rzeczywistych obrazów strefy na elektroniczne sygnały wideo na pierwszym poziomie rozdzielczości. System zawiera środki do wyodrębniania ruchu osoby lub osób znajdujących się w strefie z sygnału wideo pochodzącego z co najmniej jednej kamery wideo. Sygnały wideo wyodrębnionych ruchów danej osoby są elektronicznie porównywane ze znanymi cechami charakterystycznymi ruchów wskazujących, że dana osoba ma zamiary przestępcze. Następnie określa się poziom zamiarów przestępczych danej osoby lub osób i generowany jest odpowiedni sygnał alarmowy. Podczas prób analizy wzorców zachowań w celu ustalenia zamiarów przestępczych bardzo trudno jest uniknąć fałszywych alarmów. Fałszywe alarmy mają tendencję do „uczenia” osób korzystających z systemu nadzoru, aby nie reagowały na sytuacje uznane za niebezpieczne.

[0003] W dokumencie patentowym USA nr US9524426 ujawniono system monitorowania człowieka zawierający pewną liczbę kamer i procesor wizualny. Pewna liczba kamer jest rozmieszczona w obszarze roboczym, przy czym każda kamera jest przystosowana do przechwytywania sygnału wideo zawierającego pewną liczbę klatek obrazu,

przy czym ta pewna liczba klatek obrazu jest zsynchronizowana w czasie pomiędzy odpowiednimi kamerami. Procesor wizualny jest przystosowany do odbioru pewnej liczby klatek obrazu z pewnej liczby wizyjnych urządzeń obrazujących i wykrywania obecności człowieka na podstawie co najmniej jednej z pewnej liczby klatek obrazu za pomocą dopasowywania wzorców realizowanego na obrazie wejściowym. Obrazem wejściowym dla dopasowywania wzorców jest wyznaczona oknem przesuwnym część klatki obrazu, wyrównana względem skorygowanego układu współrzędnych w taki sposób, że oś pionowa w obszarze roboczym jest w linii z pionową osią obrazu wejściowego. W publikacji WO246662 A1 międzynarodowego zgłoszenia patentowego ujawniono sposób nadzoru stref niebezpiecznych, zwłaszcza peronów na stacjach, oparty na określonej geometrii detekcji dzielącej nadzorowany obszar na pary trapezów, przy czym jeden trapez jest wykorzystywany do detekcji, a drugi do kwalifikacji detekcji. Geometria detekcji jest połączona z etapami czułej detekcji i kwalifikacji na podstawie analizy obrazu wideo. Aby w mierzalny ilościowo sposób zwiększyć czułość na zmiany optyczne obiektu, zmiany wartości szarości pikseli pogrupowano w podmacierze o wymiarach NxN pikseli. Po eksperymentach z większą zmiennością długości macierzy N analiza wykazała, że powyżej 5 pikseli wyniki pozostają niewrażliwe na zmiany stochastyczne. Ze względu na pomiary geometryczne zachowano podmacierz obrazu o wymiarach 7 na 7 pikseli, grupującą 49 pikseli. Rozmiar takiej podmacierzy jest w każdym przypadku znacznie poniżej progu wykrywalności wynoszącego 30 cm (np. w skrajnym przypadku najbardziej odległego trapezu), więc rozdzielczość geometryczna jest wystarczająca we wszystkich geometriach. Zastosowanie splotowych sieci neuronowych do wykrywania anomalii w sygnałach wideo zostało ujawnione w artykule M. Sabokrou, M. Fayyaza, M. Fathy, Z. Moayeda, Deep-anomaly: Fully convolutional neural network for fast anomaly Detection in crowded scenes.

[0004] Celem wynalazku jest zapewnienie niezawodnych, automatycznych środków wspomagających osobę nadzorującą w wykrywaniu

niebezpiecznych sytuacji z udziałem niebezpiecznych przedmiotów, poprawiających czas reakcji tej osoby. Cel ten osiąga się za pomocą realizowanego komputerowo sposobu przetwarzania sekwencji klatek w celu wygenerowania alarmu w przypadku użycia niebezpiecznego obiektu, obejmującego etap pozyskania klatki, etap podziału na podobszary, etap identyfikacji obiektu i etap podejmowania decyzji, w którym to sposobie, jeśli decyzja podjęta na etapie podejmowania decyzji jest pozytywna, realizowany jest etap generowania alarmu. Etap podziału na podobszary obejmuje podział klatki na podobszary, które nakładają się na siebie i mają wymiary co najmniej 100x100 pikseli - powyżej tej wartości rozpoznawanie niebezpiecznych obiektów przez sieć neuronową jest bardziej niezawodne. Stosowanie nakładających się podobszarów zmniejsza ryzyko niewykrycia obiektu znajdującego się na granicy podobszarów. Etap identyfikacji obiektu wykonywany jest dla kolejnych podobszarów i obejmuje przetwarzanie klatki za pomocą głębokiej sieci neuronowej nauczonej wykrywania niebezpiecznego obiektu i zwraca prawdopodobieństwo wykrycia dla podobszarów klatki w postaci mapy prawdopodobieństwa dla przetworzonej klatki. Zastosowanie głębokiej sieci neuronowej umożliwia skuteczne wykrywanie i niezawodne uczenie. Decyzja podjęta na etapie podejmowania decyzji zależy od co najmniej pierwszego kryterium i drugiego kryterium zastosowanych do kolejnych klatek i odpowiadających im map prawdopodobieństwa. Pierwsze kryterium obejmuje prawdopodobieństwo wykrycia w zadanej liczbie N kolejnych klatek przekraczające zadaną wartość p_{min}. Drugie kryterium obejmuje minimalną odległość pomiędzy podobszarami, dla których w kolejnych klatkach zaobserwowano prawdopodobieństwo wykrycia przekraczające zadaną wartość p_{min}, mniejszą od zadanej wartości d_{max}. Wykazano, że jeśli alarmy pojawiają się zbyt często, osoby nadzorujące systemy monitorowania mają tendencję do wolniejszego reagowania lub ignorowania alarmów. Za rozsądne maksimum uznano jeden fałszywy alarm dziennie. Dlatego też selektywność jest

kluczowa dla systemu rozpoznawania niebezpiecznych obiektów. Połączenie kryteriów w sposobie według wynalazku umożliwia uzyskanie selektywności wystarczającej, aby liczba fałszywych alarmów nie przekraczała jednego dziennie.

[0005] Korzystnie wartość d_{max} jest zadawana podczas kalibracji systemu lub jest wynikiem filtracji mapy prawdopodobieństwa filtrem Kalmana.

[0006] Korzystnie klatki i odpowiadające im mapy prawdopodobieństwa są przechowywane w bazie danych. Przechowywanie klatek i map prawdopodobieństwa umożliwia adaptację procesu uczenia sieci neuronowej do konkretnych sytuacji.

[0007] Pozyskiwanie klatki korzystnie obejmuje wykrywanie i usuwanie tła, co oszczędza moc obliczeniową i dodatkowo zmniejsza liczbę fałszywych alarmów.

[0008] Przed użyciem na etapie podejmowania decyzji mapa prawdopodobieństwa jest korzystnie poddawana filtracji górnoprzepustowej. W rezultacie wykrywane są tylko sytuacje dynamiczne i unika się fałszywych alarmów, a system pozostaje wrażliwy na typowe niebezpieczne sytuacje.

[0009] Sposób według wynalazku korzystnie obejmuje ponadto etap równoważenia obciążenia, w którym określa się obciążenie obliczeniowe danego etapu i w razie potrzeby odrzuca się niektóre klatki wejściowe.

[0010] Cel wynalazku jest ponadto osiągany poprzez zapewnienie urządzenia do cyfrowego przetwarzania klatek, posiadającego pamięć przystosowanego do odbierania sekwencji klatek, oraz środki do generowania alarmu, które jest przystosowane do realizacji sposobu według wynalazku.

[0011] Urządzenie jest korzystnie przystosowane do równoległego odbioru co najmniej dwóch sekwencji klatek i ich równoległego przetwarzania.

[0012] Cel wynalazku jest ponadto osiągany poprzez zapewnienie produktu w postaci programu komputerowego zawierającego zestaw

instrukcji powodujących, że urządzenie do przetwarzania klatek wykonuje sposób według wynalazku.

[0013] Cel wynalazku jest ponadto osiągany poprzez zapewnienie systemu monitorowania posiadającego co najmniej jedną kamerę, która za pomocą środków komunikacji dostarcza klatki do konsoli posiadającej wyświetlacz w centrum monitorowania, w którym to systemie klatki z kamery są przetwarzane za pomocą urządzenia do cyfrowego przetwarzania klatek według wynalazku, realizującego sposób według wynalazku. W takiej konfiguracji wielkość systemu nie jest ograniczona mocą obliczeniową, nawet przy dużej liczbie kamer.

[0014] System monitorowania korzystnie ma co najmniej dwie kamery.

[0015] Korzystnie, każda z co najmniej dwóch kamer jest podłączona do osobnego urządzenia do cyfrowego przetwarzania klatek. Alternatywnie obie z co najmniej dwóch kamery są podłączone do tego samego urządzenia do cyfrowego przetwarzania klatek.

[0016] Środkiem komunikacji korzystnie jest sieć telekomunikacyjna. Zastosowanie sieci TCP/IP sprawia, że system jest bardzo elastyczny i łatwy w instalacji w typowych obiektach.

[0017] Wynalazek opisano szczegółowo poniżej w odniesieniu do figur rysunku, na którym:

fig. 1 przedstawia sieć działań przykładu wykonania sposobu według wynalazku;

fig. 2 przedstawia schemat blokowy urządzenia do cyfrowego przetwarzania klatek według wynalazku;

fig. 3a oraz fig. 3b ukazują konwersję obrazu na podobszary i mapę prawdopodobieństwa;

fig. 4 przedstawia sekwencję obrazów z poruszającą się bronią i wynikową mapą prawdopodobieństwa;

fig. 5 przedstawia schemat blokowy przykładu realizacji głębokiej sieci neuronowej;

fig. 6 przedstawia schemat podejmowania decyzji;

fig. 7 przedstawia system monitorowania z pewną liczbą kamer i urządzeń do przetwarzania klatek według wynalazku, podłączanych do obsługiwanego przez człowieka centrum monitorowania.

[0018] Na fig. 1 przedstawiono sieć działań przykładu wykonania sposobu według wynalazku. Sposób działa na cyfrowym urządzeniu przetwarzającym z zaimplementowanymi modułami realizującymi poszczególne etapy. Sposób obejmuje pozyskanie klatki 11, podczas której klatka jest pobierana z zewnętrznego urządzenia obrazującego, kamery CCD. Na tym etapie klatka jest poddawana przetwarzaniu wstępnemu. Następnie realizowany jest etap podziału na podobszary 12. Klatka jest dzielona na pewną liczbę zachodzących na siebie podobszarów, korzystnie prostokątów o krótszym boku co najmniej 100 pikseli. Podobszary 32 są wykorzystywane podczas identyfikacji obiektu 13. Każdy podobszar 32 jest przetwarzany niezależnie przy użyciu głębokiej sieci neuronowej przeszkolonej do wykrywania niebezpiecznego obiektu – ręki niosącej broń. Wynikiem działania sieci neuronowej jest prawdopodobieństwo wykrycia dla każdego podobszaru. Prawdopodobieństwa tworzą mapę prawdopodobieństwa, która jest wykorzystywana w kolejnym etapie podejmowania decyzji. Decyzja bazuje na wynikach detekcji dla 8 kolejnych klatek. Stosowane są dwa kryteria. Pierwsze kryterium obejmuje prawdopodobieństwo wykrycia w określonej liczbie 8 kolejnych klatek wyższe od zadanej wartości progowej p_{min}. Wartość progowa p_{min} wybierana jest z zakresu 0,8 ÷ 1, w zależności od rozpoznawanego obiektu. Wartość optymalna dla pistoletu mieści się w przedziale od 0,95 do 0,98. Dobre wyniki uzyskano dla 0,96. Drugie kryterium obejmuje minimalną odległości pomiędzy podobszarami, dla których w kolejnych klatkach zaobserwowano prawdopodobieństwo wykrycia przekraczające zadaną wartość p_{min} mniejszą od zadanej wartości d_{max}, która korzystnie mieści się w zakresie od 0,3 do 1 krótszego boku podobszaru. Pozwala to na zmniejszenie liczby fałszywych alarmów. Wartość d_{max} szacuje się za pomocą

algorytmu genetycznego według: Mitchell, Melanie (1996). An Introduction to Genetic Algorithms. Cambridge, MA: MIT Press. ISBN 9780585030944. Funkcja przystosowania algorytmu bazuje na następujących czynnikach: maksymalizacja współczynnika True Positive (TP), maksymalizacja klatek TP dla każdego wykrytego zdarzenia oraz minimalizacja wykryć fałszywie dodatnich (False Positive). Suboptymalna wartość d_{max} jest wybierana zgodnie ze znaczeniem Pareto.

[0019] Jeżeli oba kryteria są spełnione, wykonywany jest etap generowania alarmu 16. Jest to alarm dźwiękowy, alarm wizualny, e-mail, wiadomości tekstowe (SMS), wiadomości wideo, wiadomości głosowe (połączenia), alerty ekranowe, interfejs API do istniejących miejscowych instalacji alarmowych, interfejs API do zdalnych instalacji alarmowych lub usług organów ścigania lub dowolna kombinacja powyższych.

[0020] Na fig. 2 przedstawiono schemat blokowy urządzenia do cyfrowego przetwarzania klatek według wynalazku. Powyższe etapy sposobu realizowane są przez odpowiednie moduły oprogramowania zaimplementowane w układzie FPGA (Virtex 7) lub GPU (GeForce GTX 970, 4 GB).

[0021] Moduł H_CORE 27 jest głównym modułem systemu odpowiedzialnym za zarządzanie pozostałymi modułami. Służy do ustawiania konfiguracji ładowanej z plików konfiguracyjnych. Plik konfiguracyjny zawiera następujące parametry w formacie Windows .ini:

- Cfile: identyfikator pliku konfiguracyjnego – ciąg znaków,
- autotest: wywołanie autotestu systemu, wartość logiczna: prawda/fałsz,
- mode (tryb akwizycji wideo), liczba całkowita:
  1 - przechwytywanie wideo z pliku,
  2 - przechwytywanie wideo z kamery IP,
- filename: nazwa pliku do przetworzenia, ciąg znaków,
- camip: numer IP kamery do podłączenia, ciąg znaków,
- camuser: login użytkownika kamery, ciąg znaków,
- campwd: hasło użytkownika kamery, ciąg znaków,
- res: rozdzielczość kamery, ciąg znaków,
- fps: liczba klatek na sekundę kamery, liczba całkowita,

- codec: kodek kamery bazujący na ustawieniach kamery, liczba całkowita,
- window: rozmiar okna sieci neuronowej, liczba całkowita,
- decay: szybkość uczenia się w tle dla przetwarzania sieci neuronowej, liczba zmiennoprzecinkowa pojedynczej precyzji (float),
- probability: p_{min}, próg minimalnego akceptowanego prawdopodobieństwa sieci neuronowej, liczba zmiennoprzecinkowa pojedynczej precyzji (float), dla pistoletu 0.96,
- checkpoint: punkt kontrolny sieci neuronowej, ciąg znaków,
- maxdetections: minimalna liczba kolejnych wykryć wymagana do rozpoznania zdarzenia, liczba całkowita, zazwyczaj 8,
- maxdistance: d_{max}, maksymalna odległość pomiędzy wykryciami w kolejnych klatkach dla zdarzenia, liczba całkowita (np. 80),
- usekalman: włączenie filtracji Kalmana, wartość logiczna, prawda/fałsz,
- dbwrite: włączenie zapisu danych do bazy danych, wartość logiczna, prawda/fałsz,
- fx: współczynnik przeskalowania osi x dla klatek przed zapisem do bazy danych, liczba zmiennoprzecinkowa pojedynczej precyzji (float), np. 0.25,
- fy: współczynnik przeskalowania osi y dla klatek przed zapisem do bazy danych, liczba zmiennoprzecinkowa pojedynczej precyzji (float), np. 0.25,
- fps: liczba klatek na sekundę sekwencji wideo przechowywanej w bazie danych, liczba całkowita, np. 30,
- visualization: włączenie wizualizacji na ekranie podczas przetwarzania, wartość logiczna, prawda/fałsz.

[0022] Konfigurację można przeładować podczas pracy systemu bez konieczności ponownego uruchamiania. Moduł H_CORE 27 uruchamia inne moduły i ustanawia komunikację między nimi za pomocą kolejek danych. Moduł H_CORE 27 zarządza innymi modułami, przy czym każdy moduł może zostać zrestartowany w przypadku awarii. Sygnał stopu odebrany przez moduł H_CORE jest rozsyłany do innych modułów w celu zatrzymania całego systemu w przypadku krytycznej awarii systemu lub gdy wymagane jest wyłączenie systemu ze względu na konserwację, co jest inicjowane przez użytkownika.

[0023] Moduł H_AQ 21 pozyskuje klatkę 31 ze źródła danych, takiego jak strumień wideo z kamery 72, adres URL lub plik lokalny.

Źródła danych można wymieniać podczas normalnej pracy bez konieczności ponownego uruchamiania. W przypadku braku obrazu wejściowego moduł H_AQ 21 czeka na sygnał wejściowy. Pozyskana klatka jest sprawdzana pod kątem zgodności z wymaganymi wymiarami w pikselach. Jeśli ten test nie powiedzie się, pozyskana klatka nie jest dalej przetwarzana. Aby zaoszczędzić moc obliczeniową, przetwarzane są tylko klatki niepowtarzające się. Klatki takie wybiera się za pomocą algorytmu detekcji ruchu bazującego na modelu mieszanin gaussowskich (GMM) [Zoran Zivkovic. 2004. Improved Adaptive Gaussian Mixture Model for Background Subtraction. In Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 2 - Volume 02 (ICPR '04), Vol. 2. IEEE Computer Society, Washington, DC, USA, 28-31. DOI: http://dx.doi.org/10.1109/ICPR.2004.479] Algorytm ten wykrywa wszelkie zmiany pomiędzy bieżącą i poprzednią klatką. W przypadku braku zmian klatka traktowana jest jako powtórzona/zduplikowana i tym samym usuwana. Klatki takie nie są zatem przetwarzane przez inne moduły. W przeciwnym przypadku klatka jest umieszczana w kolejce danych pomiędzy modułem H_AQ 21 i modułem H_NN 23.

[0024] Klatka 31 z modułu pozyskiwania klatek H_AQ 21 jest przetwarzana w module wykrywania obiektów H_NN 23, który w pierwszej kolejności dzieli klatkę na nakładające się podobszary prostokątne o krótszym boku nie mniejszym niż 100 pikseli. Skok podobszaru musi być mniejszy niż połowa jego krótszego boku. Każdy podobszar jest następnie przetwarzany przez głęboką splotową sieć neuronową wytrenowaną do wyprowadzania prawdopodobieństwa wystąpienia niebezpiecznego obiektu w danym podobszarze. Sieć składa się z banku wytrenowanych filtrów cyfrowych wytwarzających cechy obrazu oraz w pełni połączonego wielowarstwowego perceptronu na potrzeby ostatecznej klasyfikacji. W rezultacie powstaje mapa prawdopodobieństwa odpowiadająca danej klatce, wypełniona wartościami prawdopodobieństwa odpowiadającymi poszczególnym podobszarom w obrębie klatki. Zależność pomiędzy

klatką a macierzą prawdopodobieństwa pokazano schematycznie na fig. 3a i b. Na fig. 3a pokazano klatkę 31 z osobą niosącą niebezpieczny przedmiot 33. Klatka 31 jest podzielona na podobszary 32. Na fig. 3b pokazana jest odpowiadająca jej mapa prawdopodobieństwa 34. Maksymalna wartość 35 na mapie prawdopodobieństwa odpowiada podobszarowi zawierającemu niebezpieczny obiekt 33. Sekwencja klatek rzeczywistego scenariusza i odpowiadających im map prawdopodobieństwa z wartościami prawdopodobieństwami przedstawionymi w postaci intensywności pokazano na fig. 4.

[0025] Przetwarzanie podobszaru w module H_NN 23 przedstawiono na fig. 5. Podobszar wejściowy 32 pełnej klatki 31 jest podawany do splotowej sieci neuronowej, która składa się z 5 kolejnych bloków splotowych 56. Każdy z tych bloków zawiera dwie splotowe warstwy. Jedna warstwa ma trenowane filtry o rozmiarze 3x3 52 i zestaw trenowanych filtrów o rozmiarze 1x1 53. Druga warstwa to warstwa max pooling 54. Dane wyjściowe bloków splotowych są następnie przekazywane do w pełni połączonej warstwy 55, która wystawia prawdopodobieństwo, że klatka wejściowa 31 zawiera niebezpieczne narzędzie.

[0026] Aby uniknąć ciągłej błędnej klasyfikacji w niektórych podobszarach odpowiadających stacjonarnym obiektom tła przypominającym niebezpieczny obiekt, sieć neuronowa jest trenowana przy użyciu standardowego algorytmu propagacji wstecznej do rozpoznawania, mapa prawdopodobieństwa jest filtrowana przez górnoprzepustowy cyfrowy filtr czasowy. Filtr ten działa w następujący sposób. Najpierw, przy przy wykorzystaniu bufora przechowujący jedną poprzednią klatkę, szacowane jest tło prawdopodobieństwa według wzoru:

$$b(t+1) = \alpha * b(t) + (1-\alpha) * p(t),$$

gdzie p(t) jest wyjściem sieci neuronowej, a α jest parametrem kontrolującym szybkość estymacji tła lub równoważnie odpowiedź impulsową filtra cyfrowego. Następnie obliczana jest aktualizacja tła u(t) = p(t)-b(t). W tym podejściu statyczna

błędna klasyfikacja jest usuwana za pomocą operacji odejmowania. Ostatecznym wyjściem modułu wykrywania obiektów 23 H_NN są współrzędne podobszaru o najwyższej wartości u(t).

[0027] Klatka i dane wyjściowe z modułu H_NN 23 są przesyłane do modułu H_DM 24, gdzie podejmowana jest ostateczna decyzja o podniesieniu alarmu. Schemat procesu podejmowania decyzji przedstawiono na fig. 6. W etapie 61 odbierane są wyniki z modułu H_NN 23. W etapie 62 moduł H_DM 24 przyjmuje zadaną liczbę N, gdzie N mieści się w zakresie 6-10 w zależności od rodzaju niebezpiecznego narzędzia, kolejnych klatek dla zapewnienia niezawodnego alarmu w przypadku użycia broni i unikania fałszywych wykryć. Według testów przeprowadzonych z ludzkimi nadzorcami, aby nadzorcy nie byli przekonani o fałszywości wykryć, dopuszczalne jest tylko jedno fałszywe wykrycie dziennie. Do wykrywania pistoletu najlepszym wyborem okazała się liczba 8. W etapie 63 jest spawdzane, czy we wszystkich 8 kolejnych klatkach maksymalne prawdopodobieństwo wykrycia przekracza wartość progową. Dla broni krótkiej najlepszą wartością progową okazało się 0.96. Jeżeli prawdopodobieństwo przekracza wartość progową dla wszystkich 8 kolejnych klatek, wykonywany jest etap 64, w przeciwnym razie ponownie wykonywany jest etap 62. W etapie 64 sprawdzane jest, czy odległości pomiędzy kolejnymi wykryciami (maksymalne prawdopodobieństwa) są mniejsze niż d_{max}. Jeżeli tak jest, w etapie 65 zostanie wygenerowany alarm. W przeciwnym przypadku ponownie zostanie wykonany etap 62.

[0028] Dobre efekty można uzyskać także dla innej liczby kolejnych wykryć w kolejnych klatkach. Rozsądne efekty uzyskano dla liczb z zakresu od 6 do 10. Ponadto odległość pomiędzy maksimami w sąsiednich klatkach nie może przekraczać możliwości modelu człowieka. Można to uzyskać na dwa sposoby. Można je stosować niezależnie lub łączyć.

[0029] Pierwszym podejściem jest metoda filtra Kalmana z zestawem zadanych parametrów, która przewiduje współrzędne kolejnego

wykrycia. Parametry dobierane są w zależności od konkretnej instalacji, przykładowe wartości macierzy są następujące:

[[1, 0, 0, 0],
[0, 1, 0, 0]],

[[1, 0, 1, 0],
[0, 1, 0, 1],
[0, 0, 1, 0],
[0, 0, 0, 1]],

[[1, 0, 0, 0],
[0, 1, 0, 0],
[0, 0, 1, 0],
[0, 0, 0, 1]].

Macierz zawiera wagi wykorzystywane w procesie predykcji. Macierz przejścia służy do określania położenia obiektu w przyszłości [Mohinder S. Grewal and Angus P. Andrews. 2014. Kalman Filtering: Theory and Practice with MATLAB (4th ed.). Wiley-IEEE Press]. Jeśli odległość pomiędzy predykcją a bieżącym wykryciem jest mniejsza niż zadana wartość, wykrycie jest klasyfikowane jako należące do tego samego zdarzenia.

[0030] Drugie podejście bazuje na obliczonej odległości pomiędzy pozycjami kolejnych wykryć. Jeśli odległość ta jest mniejsza niż zadana wartość progowa d_{max}, wykrycie jest klasyfikowane jako należące do tego samego zdarzenia. Odległość jest obliczana jako miara euklidesowa wyrażona w pikselach. Wartość progowa d_{max} wynosi 0.3 ÷ 1 rozmiaru podobszaru.

[0031] Alarm jest podnoszony przez moduł powiadamiania użytkownika H_NT 26, gdy zdarzenie trwa dłużej niż zadana liczba klatek. Zwykle było to 8, ale okazało się, że 6–10 jest zakresem, w którym osiągana jest maksymalna dzienna liczba fałszywych wykryć. Zwiększanie tej wartości ogranicza czułość systemu, natomiast jej zmniejszanie zwiększa czułość systemu. Zwiększenie czułości skutkuje wyższym prawdopodobieństwem wykrycia kosztem większego prawdopodobieństwa fałszywego alarmu. Zmniejszenie

czułości daje odwrotny efekt. Klatka wraz z informacją o zgłoszonym alarmie umieszczana jest w kolejce danych pomiędzy modułem H_DM 24 a modułem H_DB 25.

[0032] Klatka i dane wyjściowe z modułu podejmowania decyzji H_DM 24 są wysyłane do modułu H_DB 25, w którym dane są zapisywane w bazie danych. Moduł H_DB 25 zapisuje wszystkie alarmy i sekwencje wideo, które były przyczyną tych alarmów. Moduł H_DB 26 agreguje klatki z zadanego okresu w celu zbuforowania wideo w bazie danych i w przypadku wykrycia zdarzenia i wywołania alarmu zapisania na stałe sekwencji klatek rozpoczynającej się przed zdarzeniem i kończącej się po zdarzeniu. Ostatnie zapisane wideo rozpoczyna się N sekund przed zdarzeniem i kończy N sekund później. Wartość N ustalana jest przez użytkownika, typowa wartość to 15. Zwiększenie tej wartości powoduje zapisanie w bazie danych szerszego kontekstu wykrytego zdarzenia, natomiast zmniejszenie tej wartości powoduje zmniejszenie ilości danych przechowywanych w bazie. Przechowywanie na stałe jedynie zapisu kontekstu, który był przyczyną alarmu, pozwala na przyszłe badania i rozwój systemu oraz dokumentowanie zdarzeń pomimo ograniczonej pojemności pamięci bazy danych. Moduł H_DB 25 zapewnia także wizualizację w czasie rzeczywistym. Na ekranie wyświetlane są użytkownikom klatki z zaznaczonymi wykrytymi obiektami i alarmami.

[0033] Moduł powiadamiania użytkownika H_NT 26 wykorzystuje dane zgromadzone w bazie danych i zapewnia usługę powiadamiania użytkownika. System wykorzystuje dziewięć rodzajów powiadomień: alarm dźwiękowy, alarm wizualny, e-mail, wiadomości tekstowe (SMS), wiadomości wideo, wiadomości głosowe (połączenia), alert ekranowy, API do istniejących lokalnych instalacji alarmowych, API do zdalnych instalacji alarmowych lub usług organów ścigania, lub dowolną kombinację powyższych. Proces powiadamiania wykorzystuje książkę adresową z funkcją czasu w celu powiadamiania różnych użytkowników w zależności od aktualnego czasu. API umożliwia integrację systemu z oprogramowaniem stron trzecich.

[0034] Urządzenie do przetwarzania klatek według wynalazku jest kluczowym elementem systemu monitorowania według wynalazku. Przykładowy schemat blokowy systemu pokazano na fig. 7. W takim systemie kamery połączone są z centrum monitorowania 74 za pośrednictwem sieci telekomunikacyjnej. Klatki z kamery są przetwarzane za pomocą urządzeń do przetwarzania klatek według wynalazku przed dostarczeniem do centrum monitorowania 74 i pokazaniem osobie nadzorującej na ekranie konsoli 75. Alarmy generowane przez urządzenia do przetwarzania klatek są pokazywane osobie nadzorującej jednocześnie z niebezpiecznym przedmiotem. Informacje dotyczące podobszaru zawierającego niebezpieczny obiekt służą do zwrócenia uwagi osoby nadzorującej na konkretną część kadru. Jest ona podświetlona lub oznaczona okręgiem w wyświetlanej sekwencji klatek. Urządzenia do przetwarzania klatek mogą współpracować indywidualnie z pojedynczą kamerą lub mogą otrzymywać wiele strumieni klatek z wielu kamer przetwarzanych równolegle. Właściwość ta ułatwia rozbudowę systemu monitoringu o dodatkowe kamery.

[0035] Możliwe jest zastosowanie pojedynczego centrum monitorowania 74 z obsługiwaną przez człowieka konsolą 75 z systemami zdalnego nadzoru 71a, 71b zlokalizowanymi w różnych, odseparowanych przestrzennie obiektach. Takiemu przykładowemu systemowi odpowiada schemat blokowy przedstawiony na fig. 7.

[0036] Przykładowo, każdy z systemów zdalnego nadzoru 71a, 71b ma dwie kamery 72a, 72b. W pierwszym systemie zdalnego nadzoru 71a dwie kamery 72a są bezpośrednio podłączone kablem z tym samym urządzeniem do cyfrowego przetwarzania klatek 73a, znajdującym się w tym samym obiekcie, co system zdalnego nadzoru 71a. W drugim systemie zdalnego nadzoru 71b dwie kamery 72b są połączone za pośrednictwem cyfrowej sieci komunikacyjnej z urządzeniem do cyfrowego przetwarzania klatek 73b, znajdującym się w centrum monitorowania 74 z konsolą 75 wyposażoną w ekran i obsługiwaną przez osobę nadzorującą.

[0037] System monitorowania według wynalazku jest modułowy. Dowolne elementy można łączyć za pomocą kabla, światłowodów, przewodowych lub bezprzewodowych sieci telekomunikacyjnych, w tym WiFi, LTE, Ethernet i innych. Zastosowanie wielu urządzeń do przetwarzania klatek umożliwia rozbudowę systemu, nieograniczoną przez moc obliczeniową obsługiwanej przez człowieka konsoli. Każda zwykła lub przemysłowa kamera może być rutynowo wprowadzona do systemu przez osobę wykwalifikowaną w tej dziedzinie.

[0038] Urządzenie do przetwarzania klatek według wynalazku może zostać zastosowane w istniejącym systemie, połączone z jedną lub większą liczbą kamer i centrum monitorowania. Takie późniejsze podłączenie jest szczególnie łatwe w przypadku istniejących sieci telekomunikacyjnych, które zazwyczaj są zapewnione w monitorowanych obiektach. Można do tego celu wykorzystać także radiowe sieci telefonii komórkowej.

[0039] Należy zauważyć, że opisane powyżej przykłady wykonania wynalazku mają na celu jedynie zilustrowanie, a nie ograniczenie wynalazku, który jest zdefiniowany w załączonych zastrzeżeniach. Specjaliści w dziedzinie będą w stanie zaprojektować wiele alternatywnych przykładów wykonania. Wynalazek może zostać zrealizowany w programie komputerowym zawierającym co najmniej fragmenty kodu, które realizują etapy sposobu według wynalazku po uruchomianiu na programowalnym urządzeniu do przetwarzania klatek, takim jak system komputerowy, procesor sygnałowy, układ FPGA, lub umożliwiają wspomnianemu urządzeniu do przetwarzania klatek wykonywanie funkcji urządzenia według wynalazku. Program komputerowy może na przykład zawierać jeden lub większą liczbę z następujących elementów: podprogram, funkcja, procedura, metoda obiektowa, implementacja obiektu, aplikacja wykonywalna, aplet, servlet, kod źródłowy, kod obiektowy, biblioteka współdzielona /biblioteka dynamiczna i/lub inna sekwencja instrukcji przeznaczona do wykonania w systemie komputerowym. Program komputerowy może być dostarczony na nośniku danych, takim jak dysk

optyczny, dysk twardy, dysk magnetyczny lub dysk półprzewodnikowy, zawierającym dane do załadowania do pamięci systemu komputerowego, przy czym dane te reprezentują program komputerowy. Nośnik danych może zatem być materialnym nośnikiem danych. Nośnikiem danych może być łączem danych, takim jak kabel telefoniczny lub kabel sieciowy.

[0040] W załączonych zastrzeżeniach patentowych użycie czasownika „zawierać/obejmować” i jego odmian nie wyklucza obecności elementów lub etapów innych niż te wymienione w zastrzeżeniu. Przedimek „a” lub „an” poprzedzający element nie wyklucza obecności wielu takich elementów. Wynalazek może być realizowany za pomocą sprzętu składającego się z kilku odrębnych elementów oraz za pomocą odpowiednio zaprogramowanego komputera. W zastrzeżeniu dotyczącym urządzenia wymieniającym kilka środków, kilka z tych środków może być zawartych w jednym i tym samym elemencie sprzętu. Sam fakt, że pewne środki są przywoływane we wzajemnie różnych zastrzeżeniach zależnych nie oznacza, że połączenie tych środków nie może być korzystnie wykorzystane.

Zastrzeżenia patentowe

1. Realizowany komputerowo sposób przetwarzania sekwencji klatek dla wygenerowania alarmu w przypadku użycia niebezpiecznego obiektu, obejmujący etap pozyskania klatki (11), etap podziału na podobszary (12), etap identyfikacji obiektu (13) i etap podejmowania decyzji (14), w którym, jeśli decyzja podjęta w etapie podejmowania decyzji (14) jest pozytywna, realizowany jest etap generowania alarmu (16), **znamienny tym, że**

etap podziału na podobszary (12) obejmuje podział klatki na podobszary, które nakładają się na siebie i mają wymiary co najmniej 100x100 pikseli,

etap identyfikacji obiektu (13) wykonywany jest dla kolejnych podobszarów i obejmuje przetwarzanie klatki za pomocą głębokiej sieci neuronowej wytrenowanej do wykrywania niebezpiecznych obiektów i zwraca prawdopodobieństwa wykrycia dla podobszarów klatki w postaci mapy prawdopodobieństwa dla danej klatki,

decyzja podjęta w etapie podejmowania decyzji (14) zależy od kombinacji co najmniej pierwszego kryterium i drugiego kryterium zastosowanych do kolejnych klatek i odpowiadających im map prawdopodobieństwa,

przy czym pierwsze kryterium obejmuje przekroczenie zadanej wartości p_{min} przez prawdopodobieństwo wykrycia, kolejno w zadanej liczbie N klatek,

a drugie kryterium obejmuje stwierdzenie, czy odległość pomiędzy podobszarami, dla których w kolejnych klatkach zaobserwowano prawdopodobieństwo wykrycia przekraczające zadaną wartość p_{min} jest mniejsza od zadanej wartości d_{max}.

2. Sposób według zastrz. 1, w którym wartość d_{max} jest określana podczas kalibracji systemu.

3. Sposób według zastrz. 1, w którym wartość d_{max} jest wynikiem filtracji mapy prawdopodobieństwa filtrem Kalmana.

4. Sposób według zastrz. 1 albo 2 albo 3, w którym klatki i odpowiadające im mapy prawdopodobieństwa są przechowywane w bazie danych.

5. Sposób według któregokolwiek z zastrz. od 1 do 4, w którym pozyskanie klatki obejmuje wykrywanie i usuwanie tła.

6. Sposób według któregokolwiek z zastrz. od 1 do 4, w którym mapa prawdopodobieństwa jest poddawana filtracji górnoprzepustowej przed wykorzystaniem w etapie podejmowania decyzji (14).

7. Sposób według któregokolwiek z zastrz. od 1 do 5, obejmujący ponadto etap równoważenia obciążenia, w którym określa się obciążenie obliczeniowe i w razie potrzeby odrzuca się niektóre klatki wejściowe.

8. Urządzenie do cyfrowego przetwarzania klatek posiadające pamięć i przystosowane do odbierania sekwencji klatek oraz środki do generowania alarmu, **znamienne tym, że** jest ono przystosowane do wykonywania sposobu według któregokolwiek z zastrz. 1 do 7.

9. Urządzenie do cyfrowego przetwarzania klatek według zastrz. 8 przystosowane do równoległego odbioru co najmniej dwóch sekwencji klatek i równoległego ich przetwarzania.

10. Produkt w postaci programu komputerowego przystosowany do sterowania urządzeniem do przetwarzania klatek, **znamienny tym, że** zawiera zestaw instrukcji powodujących wykonanie przez to urządzenie sposobu według któregokolwiek z zastrz. 1 do 7.

11. System monitorowania posiadający co najmniej jedną kamerę (72,72a,72b), która za pomocą środków komunikacji dostarcza

klatki do konsoli (75) posiadającej wyświetlacz w centrum monitorowania (74), **znamienny tym, że** zawiera on urządzenie do cyfrowego przetwarzania klatek (73a, 73b) według zastrz. 8 albo 9, do którego podłączona jest kamera (72, 72a, 72b), przy czym wspomniane urządzenie do przetwarzania (73a, 73b) realizuje sposób według dowolnego z zastrz. 1 do 7 w celu przetwarzania klatek z kamery.

12. System monitorowania według zastrz. 11, posiadający co najmniej dwie kamery (72a, 72b).

13. System monitorowania według zastrz. 12, w którym każda z przynajmniej dwóch kamer (72a, 72b) jest podłączona do osobnego urządzenia do cyfrowego przetwarzania klatek.

14. System monitorowania według zastrz. 12, w którym obie z co najmniej dwóch kamer (72a) są podłączone do tego samego urządzenia do cyfrowego przetwarzania klatek (73a).

15. System monitorowania według dowolnego z zastrz. 11-14, w którym środkiem komunikacji jest sieć telekomunikacyjna.

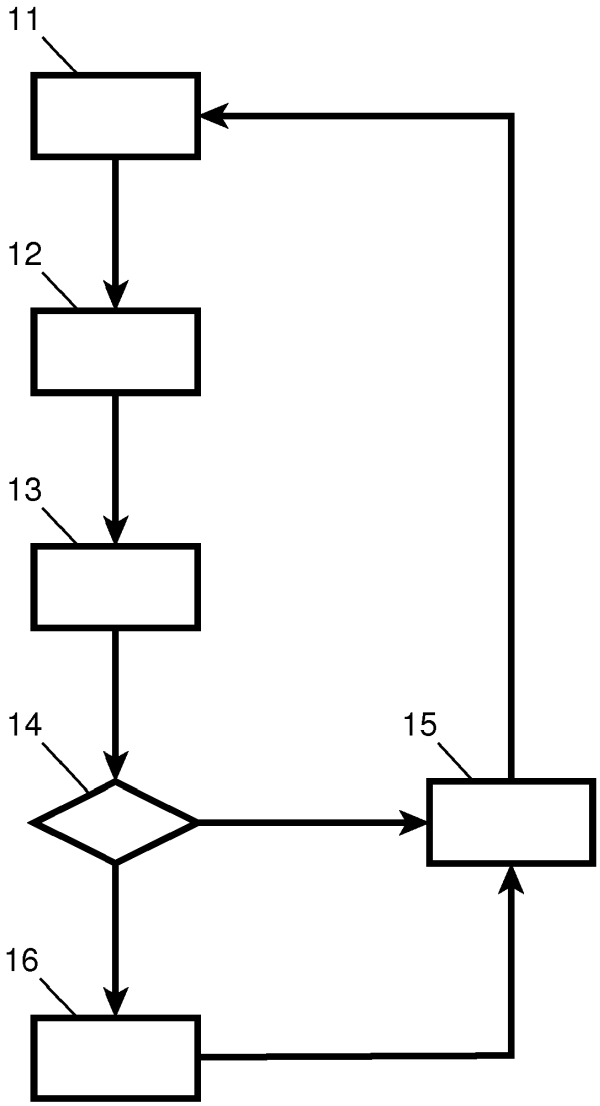

Fig. 1

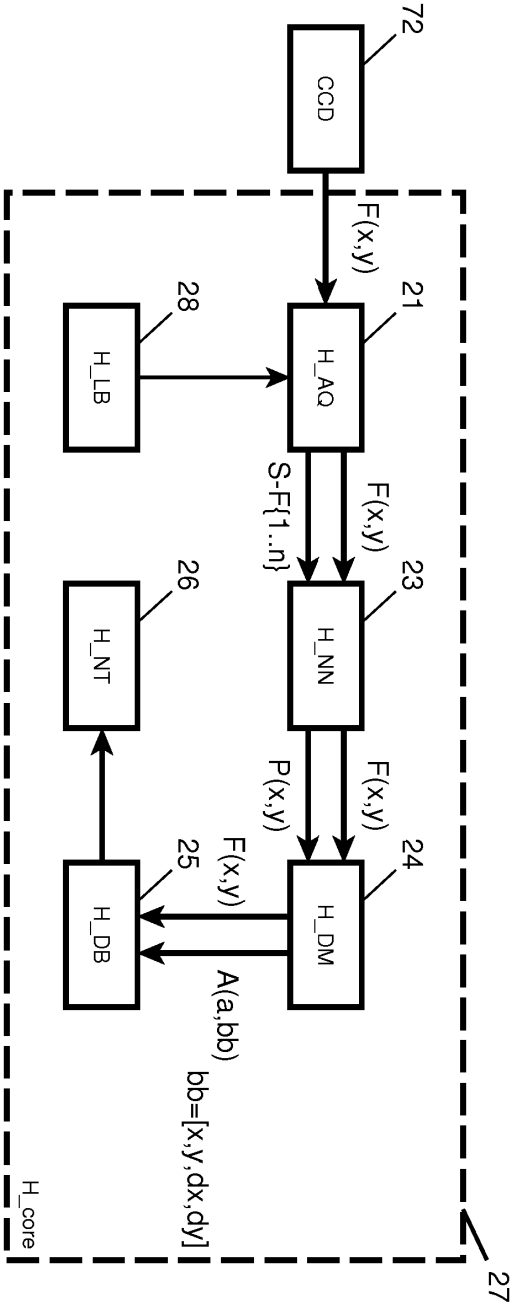
72
CCD
F(x,y)
21
H_AQ
28
H_LB
S-F{1..n}
F(x,y)
23
H_NN
26
H_NT
P(x,y)
F(x,y)
24
H_DM
F(x,y)
25
H_DB
A(a,bb)
bb=[x,y,dx,dy]
H_core
27

Fig. 2

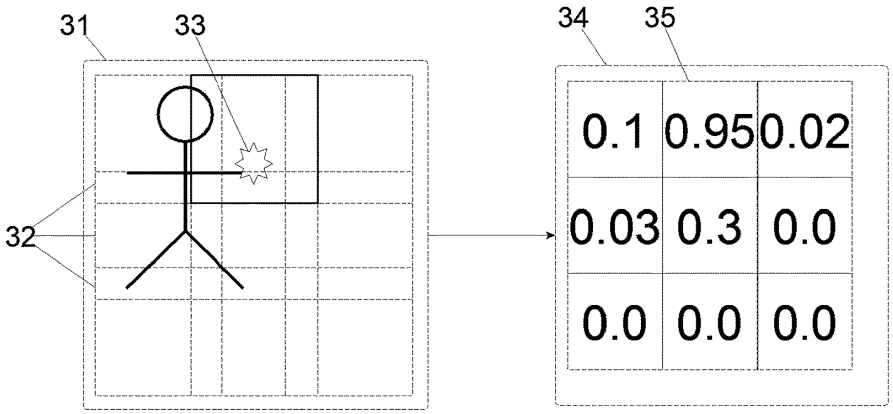

Fig. 3a Fig. 3b

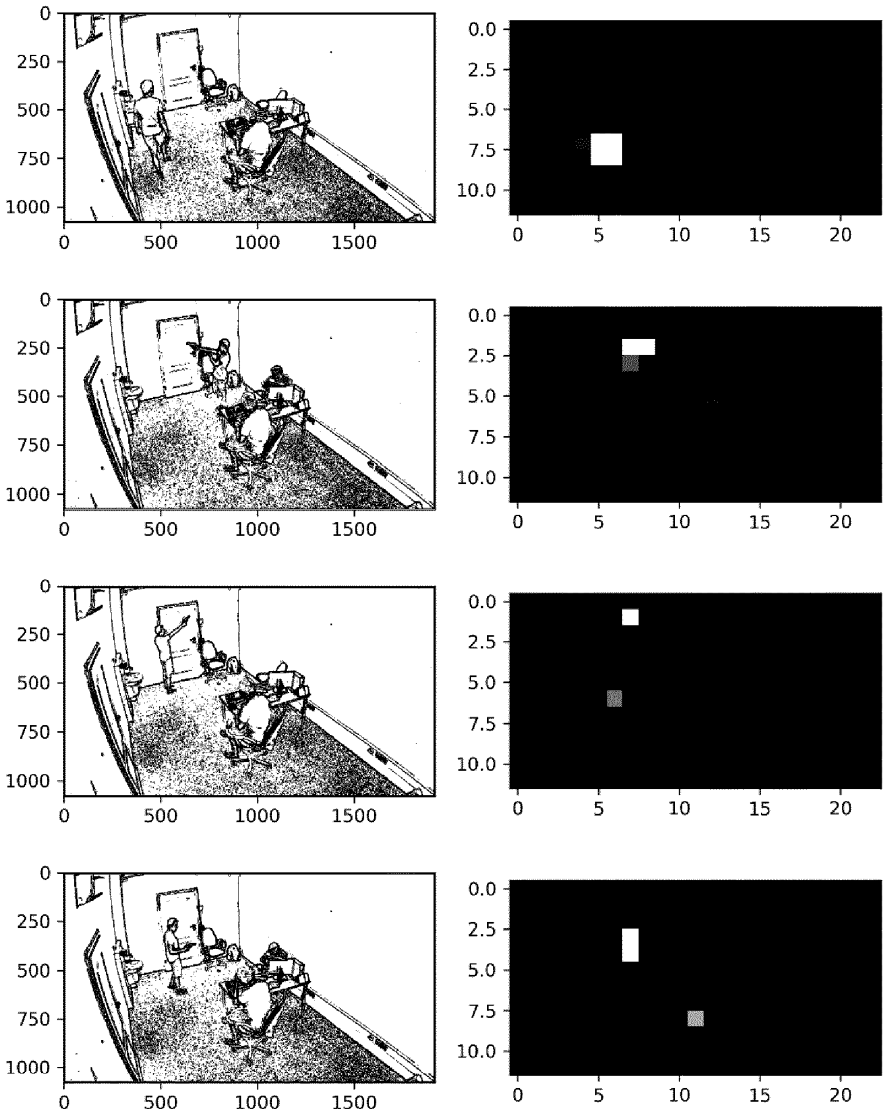

Fig. 4

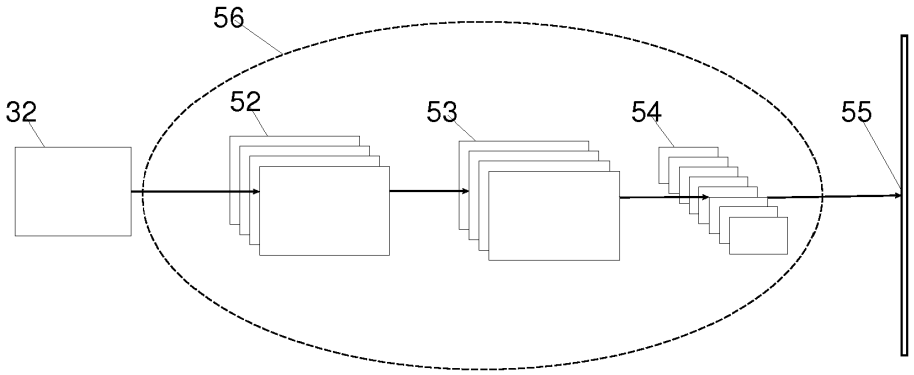

Fig. 5

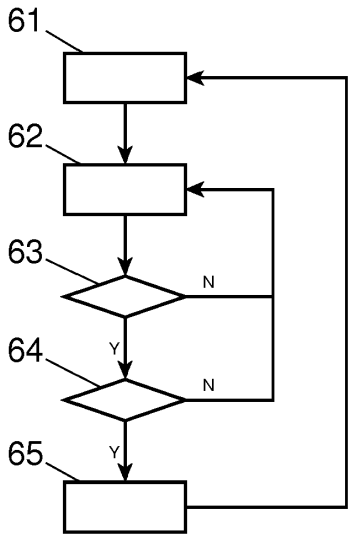

Fig. 6

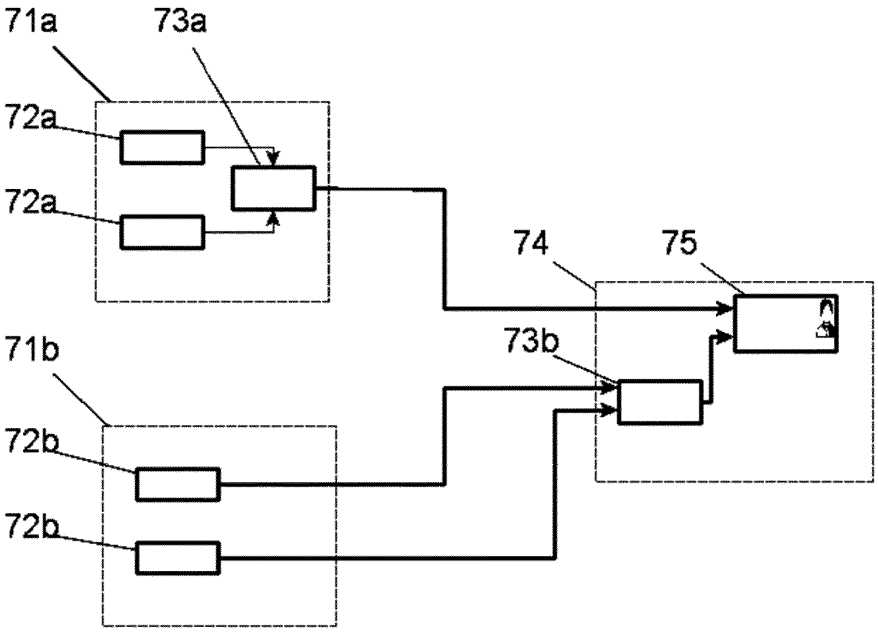

Fig. 7