Potrzeba sięgania do techniki automatycznego rozumienia jako bazy systemów zabezpieczania inteligentnych budynków

Ryszard Tadeusiewicz

Streszczenie: W artykule przedstawiono ogólną koncepcję systemu automatycznego rozumienia obrazu, który może polepszyć stan bezpieczeństwa inteligentnych budynków. Uzasadniono potrzebę stworzenia takiego systemu i pokazano, na czym polega różnica pomiędzy jego działaniem a funkcjonowaniem bardziej znanych systemów automatycznego rozpoznawania obrazów. Wprowadzono pojęcie zasobu wiedzy ekspertów jako klucza do automatycznego rozumienia (semantycznej analizy) obrazu oraz pojęcie rezonansu kognitywnego. Zwłaszcza to ostatnie pojęcie, będące podstawą kojarzenia strumienia danych sensorycznych z oczekiwaniami wynikającymi z wiedzy ekspertów, ma w tej pracy fundamentalne znaczenie. Artykuł nawiązuje do wcześniejszych prac autora, w których automatyczne rozumienie obrazów było wykorzystywane w odniesieniu do zadań wspomagania diagnostyki medycznej, ale uwzględnia specyfikę wynikającą z konieczności powiązania rozważanych treści z potrzebami twórców inteligentnych budynków.

Abstract: *The article presents the general concept of an automatic image understanding, which can improve the security services in intelligent buildings. It was justified by expressing the need to create such a system, and shows the difference between the operation and functioning of automatic image understanding system and better known of automatic image recognition. The article introduced the concept of resource expert knowledge as the key to the automatic understanding (semantic analysis) of the image and the concept of cognitive resonance. Especially this last concept underlying the association sensory data stream with the expectations arising from the expert knowledge, forms the job of fundamental importance. The article refers to the author's earlier works, in which the automatic image understanding idea was used in relation to the tasks aided medical diagnosis. But idea presented here takes into account the specificity resulting from the necessity of linking content considered in article with the needs of the creators of intelligent buildings.*

Wprowadzenie

Jednym z ważniejszych zadań, jakie muszą sobie stale stawiać twórcy inteligentnych budynków, jest zagwarantowanie bezpieczeństwa mieszkańcom tych budynków, a także znajdującym się w nich przedsiębiorstwom i instytucjom. Zadanie to zostanie w niniejszym artykule przeanalizowane głębiej, niż to się zwykle robi, bo obok spraw oczywistych i ogólnie znanych – takich jak rozmieszczanie kamer obserwacyjnych i organizacja centrów obserwacji i rejestracji danych z monitoringu – rozważony zostanie problem komputerowego wspomagania semantycznej analizy tych danych.

Zaczniemy od stwierdzeń na pozór oczywistych – ale porządkujących i systematyzujących fakty. Takie ogólne wprowadzenie jest potrzebne do tego, żeby omawiane w tym artykule oryginalne zagadnienia systemów automatycznego rozumienia gromadzonych danych (nazywanych też systemami semantycznej analizy obrazów i innych sygnałów) we właściwy sposób ulokować i powiązać z pozostałymi elementami systemów zabezpieczania inteligentnych budynków. Otóż zaczynając od wyartykułowania owych stwierdzeń oczywistych, przytoczymy następujące spostrzeżenia:

Żeby wykryć, a następnie unicestwić zagrożenie, potrzeba dwóch rzeczy. Po pierwsze, trzeba dysponować środkami potrzebnymi do tego, żeby zarejestrować symptomy zagrożenia. Temu celowi służą opisane w licznych publikacjach systemy rejestracji oraz przetwarzania i analizy różnych sygnałów, obrazów i innych danych. Jest to warunek konieczny, bo oczywiście brak stosownych czujników, kamer i innych odbiorników informacji byłby równoznaczny ze ślepotą.

Jednak nawet bardzo pobieżna analiza zagadnienia wskazuje, że jest to warunek niewystarczający. Doświadczenie uczy bowiem, że nawet najbogatsze nagromadzenie dowolnych danych, na przykład sygnałów ze wszystkich tych czujników i przetworników pomiarowych, obrazów czy nagrań wideo z kamer, a nawet wyników identyfikacji, lokalizacji i analizy ruchu ludzi oraz przedmiotów, będących efektem działania systemów rejestracji, przetwarzania, analizy i rozpoznawania obrazów – zdecydowanie nie wystarcza.

W artykule spróbujemy wskazać, o jakie elementy trzeba wzbogacić to tradycyjne instrumentarium systemów bezpieczeństwa, żeby uzyskać znaczące zwiększenie efektywności ich działania.

Elementy zawsze obecne w systemach zabezpieczania inteligentnych budynków – niezbędne, ale niewystarczające

Proponując innowacje w systemach zabezpieczania inteligentnych budynków, trzeba zacząć od bazy, to znaczy od elementów, które są zawsze obecne w takich systemach. Tymi elementami są składniki wyposażenia, dzięki którym możliwa jest rejestracja, filtracja i analiza obrazów pochodzących ze wszystkich podlegających nadzorowi obszarów rozważanego budynku. Dzięki radykalnemu obniżeniu kosztów kamer obserwacyjnych i urządzeń rejestrujących obrazy możliwe jest obecnie zbieranie obrazów z bardzo wielu newralgicznych punktów nadzorowanego inteligentnego budynku oraz ich rejestracja w układzie czasowym i w układzie przestrzennym. Tego rodzaju wyposażenie jest bardzo cenne przy prowadzeniu różnych analiz zdarzeń wykonywanych *post factum*. To znaczy, że gdy już miało miejsce jakieś niebezpieczne czy szkodliwe wydarzenie, to można prześledzić, jak do niego doszło, a także ewentualnie ustalić sprawców i ułatwić pracę organom ścigania. Posiadanie takiego wyposażenia pełni pewną rolę w ochronie inteligentnego budynku przed działaniami różnego rodzaju przestępców czy wandali, jednak jest to rola polegająca głównie na odstraszaniu. Obraz może być lepiej albo gorzej zarejestrowany. Jeśli jego jakość pozostawia sporo do życzenia, to może zostać poprzez filtrację pozbawiony zakłóceń, zniekształceń czy także obecności na nim niepotrzebnych składników (na przykład tła). Ten etap działania systemu przedstawiono na rysunku 1.

Rys. 1. Wstępny etap działania systemu ochrony

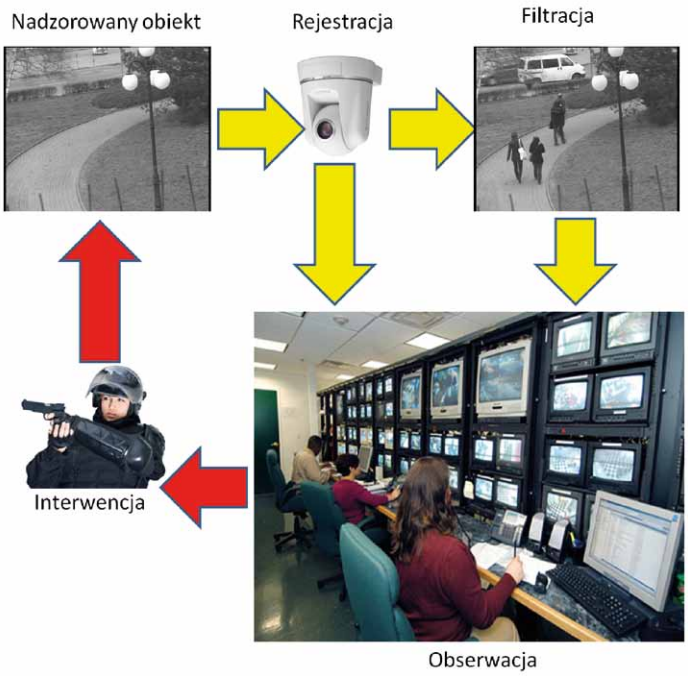
Różne metody filtracji mogą dawać bardzo istotne polepszenie jakości i czytelności rozważanego obrazu. Nie zmienia to jednak faktu, że obraz taki jest zawsze tylko zbiorem pikseli, których wartości i rozmieszczenie można dokładnie wyznaczyć, ale których

znaczenie pozostaje nieokreślone. Dlatego warunkiem koniecznym sensownego wykorzystania wszystkich urządzeń nadzoru i ochrony jest posiadanie personelu nadzoru, który obserwuje i interpretuje rejestrowane obrazy, w razie potrzeby posiłkując się także patrolami w terenie (rys. 2).

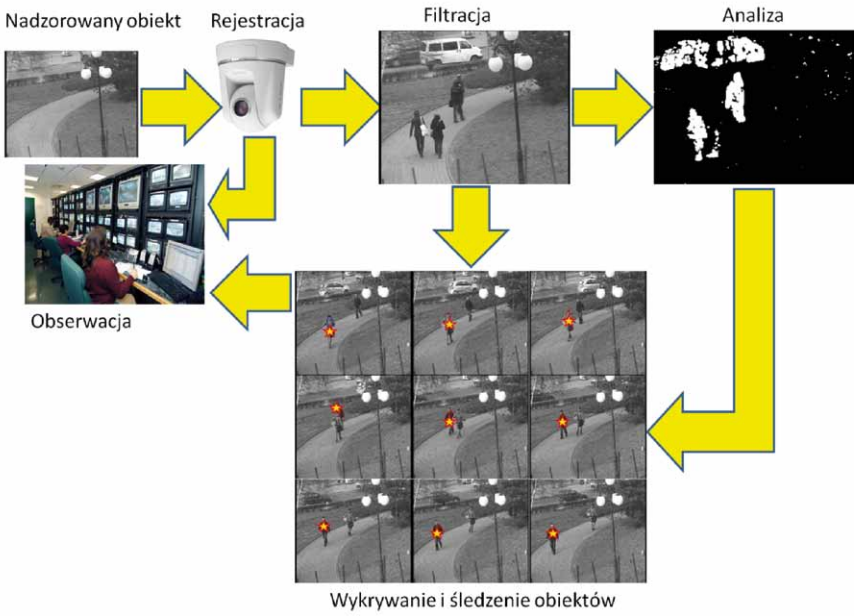
Niestety takie narzędzia nie pełnią roli zabezpieczenia aktywnego, pozwalającego w czasie rzeczywistym wykryć i zwalczyć większość aktów agresji, destrukcji czy wandalizmu w momencie, kiedy mają one miejsce. Na przeszkodzie stoi problem semantycznej (to znaczy ukierunkowanej na znaczenia, a nie na formy) interpretacji strumieni obrazów pochodzących z rosnącej liczby punktów obserwacyjnych.

Stwierdzono empirycznie, że nadzór ze strony służb ochrony budynku staje się nieskuteczny, gdy liczba obserwowanych punktów znacząco wzrasta. Dzieje się tak, ponieważ ludzie – nawet jeśli jest ich wielu – subiektywnie koncentrują uwagę na niektórych spośród obserwowanych monitorów i ignorują to, co się dzieje w polu widzenia pozostałych kamer. Oczywiście nie jest to regułą – często służby bezpieczeństwa zauważają nietypowe zachowania obserwowanych ludzi i potrafią zapobiegać zdarzeniom, które by mogły nieść zagrożenia. Ale nie jest to nigdy pewne, a gdy liczba obserwowanych punktów i obszarów wzrasta – rośnie też prawdopodobieństwo przeoczenia jakiegoś krytycznego zdarzenia uchwyconego przez jedną z kamer, ale niezauważonego przez obsługę.

Dodatkowym czynnikiem ograniczającym skuteczność typowych systemów monitoringu opartego na obsłudze ludzkiej jest monotonia pracy osób zaangażowanych do śledzenia zdarzeń i procesów na licznych monitorach. Okazuje się bowiem, że długotrwała obserwacja wielu obrazów, na których przez większość czasu nic się nie dzieje, prowadzi do sytuacji, którą w psychologii percepcji nazywa się deprywacją. Polega ona na tym, że przy długotrwałym kontakcie ze strumieniem sygnałów zmysłowych (tu – wzrokowych) niewnoszących żadnej istotnej informacji – gotowość do wychwycenia takiej istotnej informacji radykalnie maleje. Dlatego mimo posiadania odpowiednich kwalifikacji personelu i mimo dostępu do dobrej jakości danych obrazowych – niektóre sytuacje zagrożenia pozostają niewykryte. Zachodzi więc potrzeba wspomagania prac osób zaangażowanych w ochronę inteligentnych domów za pomocą odpowiednich narzędzi komputerowych. I o takich właśnie narzędziach komputerowych będzie mowa w tym artykule. Nie wystarcza przy tym najdokładniejsza nawet algorytmiczna analiza pozyskanych tym sposobem danych, ponieważ – przykładowo – symptomy zagrożenia na obrazach z kamer śledzących określony obszar czy fragment budynku są *a priori* niemożliwe do zdefiniowania.



Rys. 2. Typowy sposób wykorzystania informacji pochodzących z systemu ochrony



Rys. 3. Szkic systemu bezpieczeństwa, w którym istotną rolę odgrywa automatyczna analiza

Rozwiązanie pomocne, ale nie traktowane jeszcze jako innowacja

W rozważanym w tej pracy semantycznym systemie analizy obrazów dla innowacyjnego wspomagania systemów zabezpieczenia inteligentnych budynków nie poprzestajemy na samej tylko rejestracji i na ewentualnym przetwarzaniu obrazów (dla polepszenia ich jakości), ale dalsze etapy omawianego tu procesu komputerowego wspomagania interpretacji informacji wizyjnej w systemach bezpieczeństwa prowadzą do analizy obrazu. Różnych metod i różnych celów analizy obrazów może być dosłownie bez liku. Tutaj przykładowo na rysunku 3 pokazano możliwości automatycznego wyodrębnienia na etapie analizy sylwetek ludzkich oraz zasygnalizowano efekty działania algorytmów pozwalających śledzić ruch poszczególnych ludzi.

Systemy bezpieczeństwa wzbogacone o składnik automatycznej analizy obrazu są niewątpliwie przydatne, ale – podobnie jak systemy komputerowego przetwarzania obrazów – nie wnoszą zasadniczej poprawy w sferze jego automatycznej interpretacji. Możemy na przykład wydzielić w sposób automatyczny sylwetki ludzi znajdujących się w kadrze i możemy określić trajektorię ruchu każdego z nich, ale nie potrafimy podać ścisłych algorytmicznych metod odróżnienia zachowania złodzieja czy terrorysty od zachowania zwykłego przechodnia. Automatyczna analiza obrazu może nam na przykład pozwolić zmierzyć indywidualną szybkość ruchu każdej osoby, ale nie posunie nas ani na krok w kierunku wykrycia takiego zjawiska, jak na przykład panika. Do tego potrzebne jest zrozumienie obrazu.

Automatyczne rozumienie obrazów wprowadzone zostało przez autora i współpracowników jako nowość naukowa w 2000 roku, a w 2004 zostało dokładnie opisane w książce [12]. Jednak początkowo automatyczne rozumienie obrazów było stosowane tylko do obrazów medycznych celem ich skuteczniejszej interpretacji diagnostycznej. Dopiero poczynając od 2008 roku, zaczęto tej techniki używać do bardzo różnych celów [13, 14, 15] – i na tym opieramy się w tej pracy.

Pierwszą rzeczą, jaką trzeba wyjaśnić, jest odróżnienie innowacyjnej idei automatycznego rozumienia obrazu od pozornie podobnej, ale nie identycznej, techniki automatycznego ich rozpoznawania. Spróbujemy to prześledzić nieco dokładniej.

Automatyczne rozpoznawanie

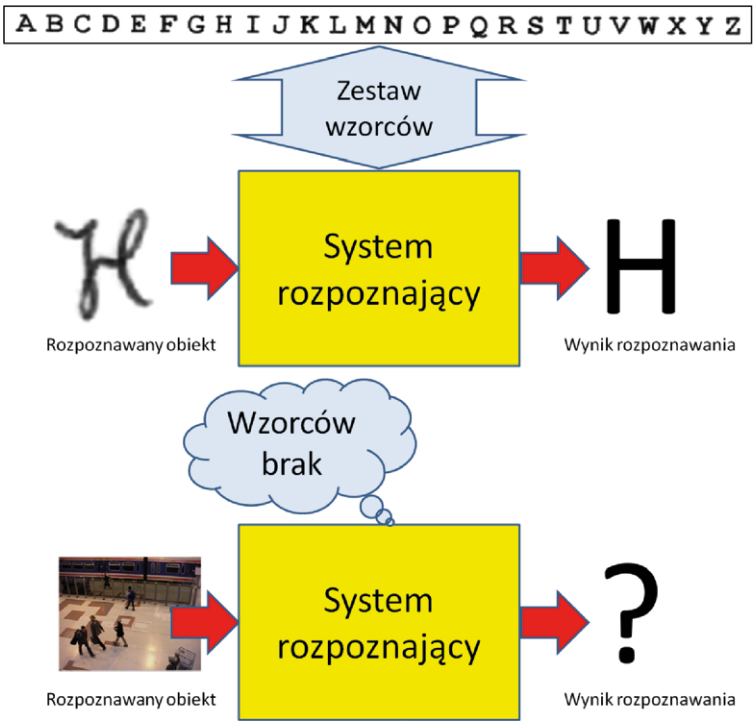
Rozwiniętą i dobrze znaną techniką, związaną z *computer vision*, jest *pattern recognition*, nazywana w Polsce niezbyt fortunnie „rozpoznawanie obrazów", podczas gdy właściwsza jest – jak się wydaje – nazwa „rozpoznawanie wzorców". Okazała się ona bardzo skuteczna w wielu zastosowaniach, na przykład w OCR (automatycznym czytaniu skanowanych tekstów drukowanych lub pisanych odręcznie), w kryminalistyce oraz w wybranych zagadnieniach diagnostyki technicznej i medycznej. Niestety zastosowanie tej techniki w zadaniach typu ochrona obiektów czy monitorowanie bezpieczeństwa napotyka na zasadnicze trudności. Wynika to z faktu, że objawy przestępczej (czy też tylko naruszającej porządek społeczny) aktywności ludzi w nadzorowanych budynkach czy obszarach specjalnego znaczenia nie mają swojego *a priori* zdefiniowanego wzorca. Istnienie takiego wzorca przyczynia się zasadniczo do skuteczności technik OCR czy do automatycznej klasyfikacji

odcisków palców. W systemach bezpieczeństwa na niepowodzenie skazana jest każda próba znalezienia jakiegoś wzorca (*pattern*) albo szablonu (*template*), podobieństwo do którego mogłoby sugerować, że wykryliśmy oto jakąś formę zagrożenia i jest powód do alarmu. Posłużmy się przykładem (rys. 4).

W typowym systemie rozpoznawania sytuacja jest prosta i oczywista: dla rozpoznawanego obiektu trzeba znaleźć wzorzec, do którego ten rozpoznawany obiekt najlepiej pasuje. Natomiast dla systemu bezpieczeństwa to nie zafunkcjonuje, nawet jeśli wyobrazimy sobie, że system wizyjny potrafi wykrywać i lokalizować sylwetki ludzkie. Sposób interpretacji obrazu, jaki w tym przypadku jest potrzebny, jest bowiem zadaniowo specyficzny. Informacją, która powinna wywołać alarm lub przynajmniej zwrócić uwagę ochrony, może być w jednych przypadkach pojawienie się ludzkiej sylwetki tam, gdzie nikogo nie powinno być – albo jej brak w miejscu, gdzie obecność człowieka (na przykład strażnika) jest wymagana. Bywają przypadki, gdy niepokojący jest fakt, że człowiek szybko się porusza (być może ucieka?), ale łatwo sobie wyobrazić sytuację, gdy zaniepokojenie budzić powinien fakt, że zauważony człowiek zatrzymał się lub porusza się wyjątkowo wolno. Może się zdarzyć, że powodem do alarmu będzie fakt, że człowiek usiadł – lub przeciwnie: że stoi, chociaż należało usiąść. Przykłady można mnożyć, nie o to jednak chodzi.

W większości wymienionych przypadków człowiek analizujący obraz mógłby (zapewne) podjąć właściwą decyzję i poprawnie zinterpretować sytuację. Z tego powodu ciężar ochrony perymetrycznej we współczesnych systemach nadzoru cały czas w dużej mierze spoczywa na pracownikach. Jest to jednak związane z szeregiem wcześniej omówionych wad, dlatego zmierzamy do stworzenia automatyzacji także procesu analizy semantycznej rozważanych obrazów, domykającego niejako system ochrony zgodnie ze schematem piramidy informacyjnej, przedstawionej na rysunku 5.

Jak stwierdzono wyżej, cechy, na podstawie których można klasyfikować czy kategoryzować sytuacje w systemach bezpieczeństwa, nie są zwykle tak oczywiste ani tak widoczne, jak (przykładowo) symptomy raka na obrazie tomograficznym wątroby badanego pacjenta albo objawy uszkodzenia turbiny parowej podczas testu diagnostycznego maszyn w elektrowni. Dlatego (poza trywialnymi przypadkami) nie da się sygnału ostrzegającego przed zagrożeniem – na przykład terrorystycznym – uzyskać drogą nawet najbardziej wyrafinowanych filtracji czy analiz danych pochodzących z czujników i przetworników,

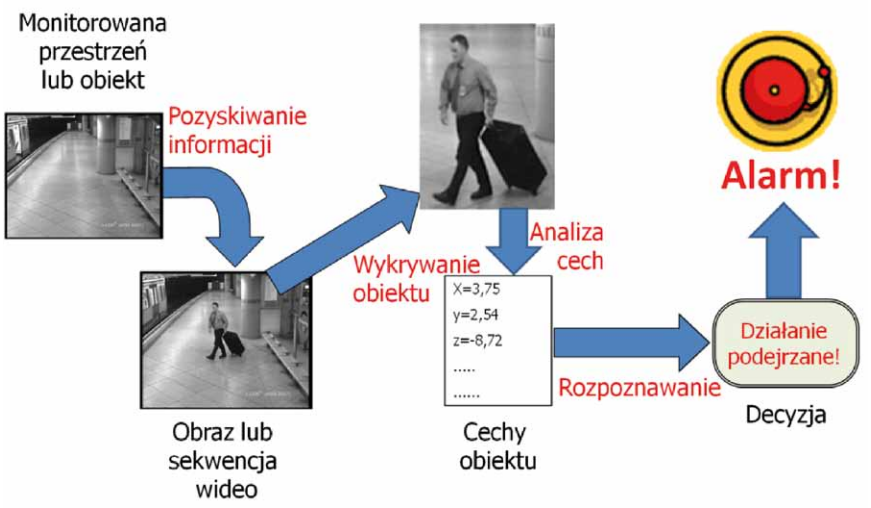


Rys. 4. Różnica pomiędzy typowym systemem rozpoznawania obrazów a systemem, który mógłby być stosowany w zadaniach ochrony, ale nie jest ze względu na brak wzorców

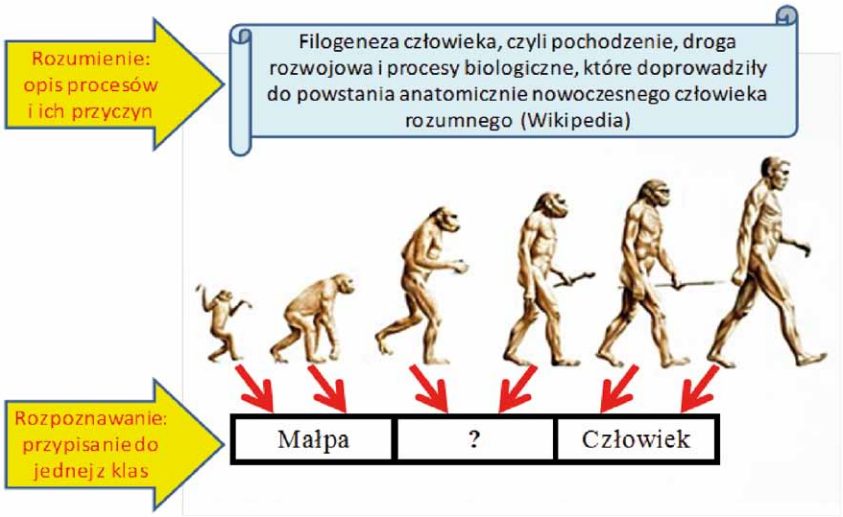
Rys. 5. Funkcje rozważanego tu systemu rozważane jako piramida oparta na surowych danych z kamer ale kończąca się automatycznym rozumieniem sytuacji, będącym podstawą ewentualnego alarmu

bo w ogólnym przypadku nie wiadomo, co podczas filtracji wydobywać, a co odrzucać, podobnie jak nie wiadomo, na czym skupić uwagę podczas analizy, a co ignorować. Schemat pokazany na rysunku 6, będący prostym przeniesieniem na grunt systemów bezpieczeństwa schematów wypracowanych w innych dziedzinach tak zwanej *computer vision* – jest po prostu nierealizowalny, bowiem bardzo trudne (wręcz niemożliwe) jest określenie *a priori* ogólnych kryteriów poprawnego lub niepokojącego zachowania obserwowanych ludzi.

Naszkicowane rozważania skłaniają do wniosku, że decyzji o tym, czy coś jest „normalne" czy „niepokojące", nie można w sposób automatyczny wyprowadzić z prostej analizy obrazu (lub sekwencji wideo), nawet połączonej z automatycznym rozpoznawaniem obiektów widocznych na obrazie czy na nagraniu. Dlatego w badaniach objętych niniejszym raportem zastosowano podejście oparte na koncepcji automatycznego rozumienia obrazów i sekwencji wideo. Podejście to stanowi logiczne domknięcie rozważanego systemu, którego ideową strukturę przedstawia w związku z tym „piramida informacyjna" przedstawiona na rysunku 5. Jak widać, przy przechodzeniu na kolejne wyższe piętra tej piramidy ilość informacji branej pod uwagę radykalnie się zmniejsza, natomiast rośnie jej wartość i przydatność z punktu widzenia celów całego projektu.



Rys. 6. Tradycyjny (nierealizowalny) sposób wykrywania niebezpiecznych zachowań i przedmiotów



Rys. 7. Ilustracja różnicy między rozpoznawaniem a rozumieniem

Automatyczne rozumienie

Skupmy się przez chwilę na odróżnieniu postulowanego automatycznego rozumienia od wzmiankowanej wyżej koncepcji automatycznego rozpoznawania (rys. 7).

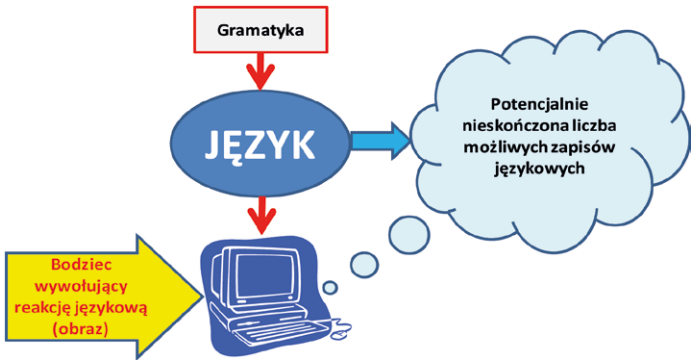
Patrząc na rysunek 7, widzimy ogólnie znaną sekwencję sylwetek istot żywych. Gdy naszym zadaniem jest rozpoznawanie, wówczas najpierw ustalamy listę możliwych klas, do których można zaliczyć analizowane obiekty. Lista taka zawsze ma skończoną, z góry określoną liczbę pozycji (wliczając w to zazwyczaj pozycję „nie wiadomo", oznaczoną na rysunku znakiem zapytania), zaś zadaniem algorytmu analizującego obraz jest stwierdzenie, do której z tych wcześniej przewidzianych klas należy zaliczyć ten czy inny konkretny obiekt. Proces rozpoznawania pokazano na rysunku 7 w jego dolnej części.

Natomiast rozumienie obrazu (osiągane przez inteligentnego człowieka, studiującego obraz, lub uzyskiwane automatycznie, do czego zmierzają badania referowane w tej pracy) oznacza wydobycie z obrazu tych wszystkich znaczeń, które są w nim *implicite* zawarte, ale nie są *explicite* widoczne (patrz rys. 6 w jego górnej części). Rozumienie dostarcza wielu wartościowych informacji i gwarantuje (w rozważanym w tej pracy zadaniu ochrony) poprawną ocenę sytuacji – wymaga jednak automatyzacji procesów kognitywnych, zachodzących oryginalnie w korze mózgowej człowieka podczas działań związanych z interpretacją rejestrowanych przez oczy obrazów, co powoduje w ogólnym przypadku spore trudności. Warto już teraz odnotować jedną z tych trudności, która będzie dalej szczegółowiej analizowana. Otóż w odróżnieniu od rozpoznawania, dla którego zbiór odpowiedzi systemu jest z góry zdeterminowany, w przypadku rozumienia sposób interpretacji obrazu jest nieprzewidywalny i z tego powodu zbiór możliwych opisów obrazu jest potencjalnie nieskończony.

Jest to poważna trudność, gdyż tę potencjalnie nieskończoną różnorodność musi wytworzyć narzędzie o bezspornie skończonych możliwościach – komputer.

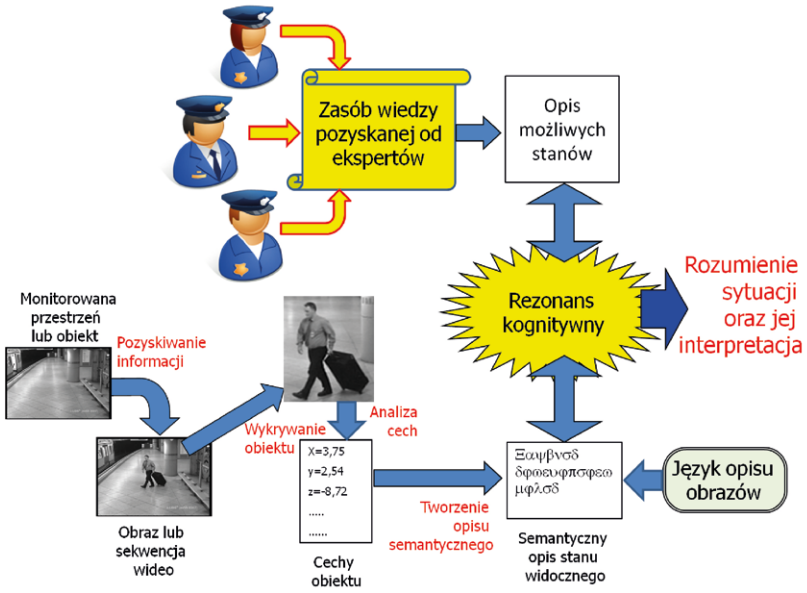
Z wcześniejszych badań prowadzonych przez autorów na nieco innym obszarze (automatycznego rozumienia obrazów medycznych) wynikał następujący wniosek:

Przy automatycznym rozumieniu obrazów pomocniczym narzędziem, którego użycie może wnieść istotny postęp w tej dziedzinie, jest lingwistyka matematyczna i obszar języków grafowych, opisujących obrazy w kategoriach pewnych wybranych elementów składowych (tak zwanych prymitywów graficznych) i ich wzajemnych relacji (rys. 8).



Rys. 8. Rola języka w rozumieniu obrazów

Wybór podejścia lingwistycznego podyktowany jest faktem, że język jest właśnie takim narzędziem, które pozwala na generowanie nieskończenie różnorodnych kombinacji, bazujących na skończonej liczbie elementów. Na przykład język polski składa się ze skończonej liczby słów i oparty jest na gramatyce mającej skończoną liczbę reguł – a jednak można w nim napisać nieskończoną liczbę artykułów, powieści, poematów, pism urzędowych itp. Również języki sztuczne (na przykład C++) cechują się tym, że mając skończoną liczbę składników oraz reguł (łatwą do opanowania przez komputerowy kompilator) – mogą służyć do wytworzenia nieograniczonej liczby programów, potencjalnie nieskończonej, po napisaniu dowolnej liczby programów zawsze możliwe jest napisanie jeszcze jednego, kolejnego.

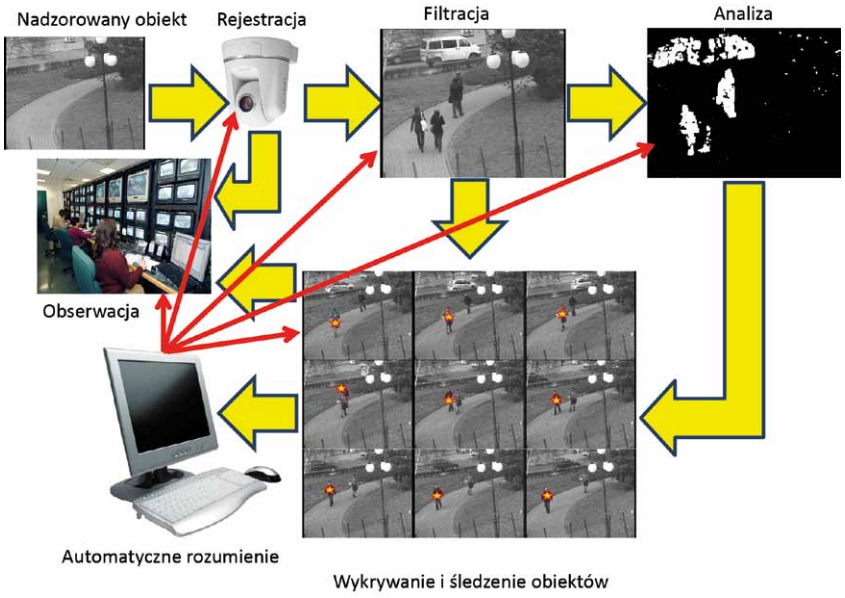


Rys. 9. Ogólny schemat systemu monitorowania wyposażonego w elementy analizy semantycznej

Obok procesu przetwarzania i analizy obrazu, ukierunkowanego na przedstawienie zawartości obrazu w postaci zapisu w odpowiednim języku grafowym, drugą cechą wyróżniającą technikę automatycznego rozumienia obrazu jest fakt, że proces wnioskowania, prowadzony w takim systemie, oparty jest na dwóch źródłach informacji (rys. 9).

Jak widać na rysunku 9, jednym z tych dwóch źródeł informacji jest analizowany obraz przedstawiający scenę, która musi być zrozumiana, żeby można było rozstrzygnąć, czy sytuacja rejestrowana przez kamery mieści się jeszcze w granicach tego, co można uznać za akceptowalne, niebudzące wątpliwości i niezmuszające do podejmowania kontroli na miejscu lub/i interwencji, czy też są podstawy do niepokoju i należy zaalarmować personel ochrony. Odpowiedni strumień danych zewnętrznych, podobnie jak w systemach tradycyjnych, zaczyna się od sensorów (na przykład kamer) i biegnie przez kolejne etapy przetwarzania, segmentacji i analizy sygnałów. Nie kończy się on jednak – jak było wyżej zapowiedziane – na identyfikacji czy kategoryzacji obiektów i przejawianych przez nie aktywności, tylko jest próbą ich scharakteryzowania za pomocą formuł specjalnie zaprojektowanego języka, o którym była mowa wyżej. Język ten aktualnie jeszcze nie istnieje, ale będzie trzeba taki język stworzyć na podstawie oceny wyników dostarczanych przez moduły przetwarzania i analizy obrazów oraz na podstawie wiedzy ekspertów – o czym będzie mowa dalej.

Drugi strumień informacji odpowiada temu, co w przypadku ludzi prowadzących obserwację tkwi w ich umysłach jako wynik odpowiedniego treningu, doświadczenia, a także po prostu ich mądrości. Ta wiedza, którą posiadają doświadczeni policjanci i strażnicy, a której nie posiadają z reguły systemy automatycznie analizujące dane z sensorów systemu monitorującego. Doświadczony policjant czy strażnik potrafi zrozumieć, co robi obserwowana osoba, ponieważ ma tę wiedzę, doświadczenie i mądrość.



Rys. 10. Umiejscowienie modułu automatycznego rozumienia obrazów w rozważanym tu systemie

Dzięki temu może odkryć w pozornie niewinnym zachowaniu obserwowanej osoby jej rzeczywiste intencje, cele i przewidywane niebezpieczne skutki działania. I odwrotnie, może zignorować zachowania pozornie niebezpieczne, prowokujące ewentualną interwencję sił porządkowych, która będzie chybiona, bo w istocie nic poważnego nie zaszło. Taki fałszywy alarm może być źródłem chorej satysfakcji dla nieodpowiedzialnych żartownisiów lub może być źródłem informacji dla rzeczywistych złodziei lub terrorystów, którzy przez takie fałszywe alarmy i pilną obserwację sposobu interwencji sił porządkowych próbują dotrzeć do nieosiągalnych dla nich w inny sposób informacji o organizacji ochrony i jej słabych punktach.

System oparty na wiedzy

Podejście omawiane w tej pracy bywa określane czasem jako oparte na wiedzy albo semantycznie zorientowane. Podejście takie określa się także niekiedy terminem kognitywistyczne, wskazując w ten sposób związek między tym podejściem a przedmiotem badań kognitywistyki jako dziedziny wiedzy o procesach poznawczych i myślowych inteligentnego człowieka.

Jeśli system automatyczny, taki, jak opisywany w tej pracy, ma inteligentnie reagować w złożonych i niejasnych sytuacjach – to trzeba go w taką wiedzę wyposażyć. Jest to możliwe, ponieważ stosowane w technice systemów ekspertowych metody pozyskiwania wiedzy od ekspertów dziedzinowych zostały już dobrze rozpracowane i wystandaryzowane. Co więcej, autor publikacji ma praktyczne doświadczenia w zakresie pozyskiwania i komputerowej implementacji wiedzy lekarzy w systemach automatycznego rozumienia wiedzy medycznej, więc można się na tym oprzeć. Niestety proces gromadzenia wiedzy jest procesem długotrwałym. W dodatku w większości przypadków wiedza, na której opierają swoje działanie (skutecznie!) pracownicy służb ochrony, jest dla nich samych wiedzą nie całkiem uświadomioną, a zawłaszcza trudną do werbalizacji. Dlatego wyposażając system w niezbędną wiedzę, trzeba opierać się zarówno na wywiadach przeprowadzanych z doświadczonymi ochroniarzami, jak i na obserwacji ich bieżącej pracy. Tego rodzaju badania są aktualnie prowadzone.

Jak widać ze schematu na rysunku 10, moduł automatycznego rozumienia dubluje niejako pracę zespołu ochroniarzy prowadzących obserwację nadzorowanego perymetru, koncentrując wysiłek na automatycznym wykrywaniu sytuacji wymagającej wzmożonej uwagi i ewentualnie także alarmu. W przypadku wykrycia w wyniku tej analizy semantycznej

jakichś sytuacji wymagających wzmożonej uwagi następuje oczywiście ostrzeżenie (zaalarmowanie) obserwatorów, ale także wynik automatycznego rozumienia obserwowanej sytuacji może skutkować zmianą sposobu rejestracji kolejnych obrazów (można zmienić częstość pobierania obrazów z określonej kamery, można zmienić jej ustawienie, wybierając inny kierunek obserwacji lub inny stopień zbliżenia (zoom). Wykrycie i semantyczne zdefiniowanie hipotetycznego zagrożenia może skutkować też zmianą metod filtracji obrazów, może skłaniać do zmiany celów i sposobów analizy obrazu bądź też uruchamiać inne, dopasowane do sytuacji, algorytmy wykrywania i śledzenia obiektów. Wszystkie te możliwości zaznaczono na rysunku 10 za pomocą czerwonych strzałek wiodących od bloku automatycznego rozumienia do odpowiednich pozostałych bloków systemu.

Jak pokazano na rysunku 9, centralnym elementem podsystemu automatycznego rozumienia musi być zasób wiedzy pozyskanej od ekspertów, którymi są w tym przypadku doświadczeni pracownicy ochrony i ewentualnie funkcjonariusze służb specjalistycznych (policjantów, strażaków, saperów itp.). Taką wiedzę trzeba będzie pozyskać i we właściwy sposób odwzorować w budowanym systemie. Nie było możliwe wykonanie tego podczas aktualnie kończonego etapu badań, będzie to więc musiało być przedmiotem dalszych prac. Tworząc odpowiednią bazę wiedzy, na której chcemy oprzeć system automatycznego rozumienia zagrożeń, trzeba będzie zwrócić uwagę na trojakiego rodzaju składniki, konieczne do pozyskania od ekspertów (rys. 11).

Pierwszym składnikiem są przesłanki. Wiedząc, jakie cechy statycznych obrazów i dynamicznych sekwencji wideo (występujące pojedynczo lub wiązane określonymi relacjami czasowymi, przestrzennymi lub przyczynowymi) są podstawą do procesu wnioskowania prowadzonego przez eksperta – możemy ustalić, jakie elementy będą musiały wchodzić w skład formuł generowanych przez wybrany język opisu obrazów dla analizowanych sytuacji. Zakładając, że wzmiankowane elementy będą pełniły rolę rzeczowników, a ustalane pomiędzy nimi relacje będą analogiem czasowników – będziemy mogli zdefiniować potrzebny język opisu obrazów. Oczywiście trzeba będzie przy tym ustalić także reguły gramatyki tego języka, wydaje się jednak, że odpowiednia powinna się tu okazać struktura gramatyki grafowej o etykietowanych krawędziach grafu, ponieważ tego typu gramatyki potwierdziły swoją użyteczność w wielu zastosowaniach.

Drugim godnym uwagi elementem, wchodzącym w skład rozważanego elementarnego składnika wiedzy eksperta, są wnioski. Są one tym elementem, z pomocą którego budować będziemy wyjście z całego podsystemu automatycznego rozumienia. Przyjmować bowiem będziemy, że automatyczne zrozumienie analizowanego obrazu lub interpretowanej sceny polegać będzie na tym, że wygenerowane zostaną automatycznie wszystkie te wnioski, jakie na temat sytuacji widocznej na obrazie lub w sekwencji wideo mógłby wyciągnąć ekspert (doświadczony ochroniarz) oglądający ten obraz lub film z maksymalną uwagą.

Rezonans kognitywny – klucz do automatycznego rozumienia

Bardzo ważnym elementem rozważanego systemu są uwidocznione na rysunku 11 reguły wnioskowania. Reguły te są wykorzystywane przez blok opisany na rysunku 9 jako rezonans kognitywny. W bloku tym generowane są automatycznie hipotezy na temat tego, jak można interpretować obraz podlegający w danym momencie analizie i opisany przez formuły języka budowanego na bazie wskazywanych przez ekspertów przesłanek. Hipotezy związane są z wnioskami podawanymi (na etapie gromadzenia wiedzy) przez ekspertów. Hipoteza może polegać na wyborze jednego z zarejestrowanych wniosków, może opierać się na równoczesnym wysunięciu kilku wniosków albo może wyrażać się poprzez zaprzeczenie wniosku (ewentualność wyrażona pewnym wnioskiem zostaje wtedy wykluczona z dalszych rozważań).

Generacja tych hipotez przebiega w sposób losowy ze zmiennym rozkładem prawdopodobieństwa. Na początku procesu adaptacji generatora hipotez do rozwiązywanego zadania przyjmowany jest pewien aprioryczny rozkład prawdopodobieństwa, wynikający z długo-

Rys. 11. Składnik wiedzy eksperta i jego elementy składowe

czasowej statystyki zdarzeń pojawiających się w ochranianym obiekcie albo przyjmowany na podstawie zewnętrznych przesłanek. Takimi zewnętrznymi przesłankami mogą być na przykład ostrzeżenia pochodzące od policji lub służb wywiadu i kontrwywiadu, uczulające ochronę budynku na specjalny rodzaj zagrożeń. Przykładowo mogą to być zapowiedzi aktu terrorystycznego albo sygnały o obecności w budynku grupy „zadymiarzy". W tych ostatnich przypadkach hipotezy zakładające możliwość pojawienia się zagrożeń związanych z tymi właśnie wybranymi i wskazanymi źródłami powinny być sprawdzane częściej niż inne – co osiąga się odpowiednio zwiększoną wartością stosownego prawdopodobieństwa.

Jak wspomniano wyżej, w trakcie funkcjonowania omawianego systemu wykorzystywane w nim rozkłady prawdopodobieństw ulegają modyfikacjom (system jest adaptacyjny!) w oparciu o ocenę skuteczności poszczególnych hipotez w budowaniu poprawnej interpretacji semantycznej zdarzeń rzeczywiście zachodzących w strzeżonym obiekcie. Jeśli w poprzednim etapie pewna hipoteza, wysunięta przez system w następstwie procesu rezonansu kognitywnego, potwierdziła się w praktyce, to prawdopodobieństwo ponownego użycia tej samej hipotezy zostaje zwiększone, a prawdopodobieństwa hipotez alternatywnych są zmniejszane w celu zachowania warunku normalizacji (suma wartości prawdopodobieństw wszystkich rozważanych hipotez musi wynosić 1).

Warto może skomentować jeszcze jedną cechę wyżej naszkicowanej koncepcji losowego generowania hipotez. Otóż preferuje ona oczywiście te hipotezy, które są najbardziej prawdopodobne, ale nie wyklucza możliwości wygenerowania hipotezy, która jest mało prawdopodobna, a jednak dla całkowitego bezpieczeństwa

powinna być także od czasu do czasu sprawdzona. W ten sposób system nie traci czujności i jest stale gotowy wykryć dowolne, nawet bardzo mało prawdopodobne i dawno niewystępujące zagrożenie, chociaż oczywiście głównie koncentruje uwagę na tych zagrożeniach, które są popularne i mogą się pojawiać najczęściej.

Na każdym etapie pracy systemu generowanych jest od kilku do kilkunastu hipotez, które następnie będą konkurowały ze sobą, wykorzystując odpowiednio gromadzoną „moc".

Proces generacji hipotez jest bowiem częścią inicjującą procedury rezonansu kognitywnego, ale nie jest częścią finalną. Dla każdej wygenerowanej hipotezy przeszukuje się bazę wiedzy i wybiera się wszystkie te elementarne składniki wiedzy ekspertów, w których ta hipoteza występowała jako wniosek. Korzystając z odpowiednich reguł, poszukuje się następnie tych przesłanek, które mogłyby rozważaną hipotezę potwierdzić, a znalazłszy je – odwołuje się do semantycznego (lingwistycznego) opisu aktualnego obrazu (czy też rozważanej sceny dynamicznej), w którym te przesłanki powinny dać się zidentyfikować. Każda przesłanka odnaleziona w opisie aktualnego obrazu będzie zwiększała „moc" rozważanej hipotezy. Każda nieznaleziona przesłanka (która powinna być, jeśli hipoteza ma być prawdziwa) – będzie tę „moc" zmniejszała. Działanie to będzie przeprowadzane równocześnie dla wszystkich rozważanych hipotez, dla wszystkich reguł, które się z nimi wiążą, oraz dla wszystkich warunków określających, kiedy takiej czy innej reguły można użyć.

Prowadząc opisany wyżej proces, obserwujemy, że „moc" pewnych hipotez maleje, innych wzrasta w niewielkim stopniu, ale może się tak zdarzyć (chociaż nie musi), że „moc" pewnej hipotezy gwałtownie wzrośnie, majoryzując wszystkie inne hipotezy. Utworzy się swoisty „pik rezonansowy", będący następstwem wzajemnego dopasowania oczekiwań wynikających z rozważanej hipotezy i rzeczywistych cech i atrybutów obrazu, wykrytych na etapie jego przetwarzania i analizy. Wystąpienie takiego rezonansu upoważnia do tego, żeby tę hipotezę, która rezonans wywołała, podać na wyjściu systemu jako domyślne (domniemane) znaczenie obrazu czy sceny, które były przedmiotem próby semantycznej interpretacji.

Zjawisko rezonansu kognitywnego jest rzadkie, więc system nie będzie zbyt skory do tego, żeby ferować wyroki na temat sposobu rozumienia obserwowanej sceny. W większości przypadków odpowiedzią systemu automatycznego rozumienia będzie… brak odpowiedzi. Jeśli jednak dojdzie do rezonansu kognitywnego, to wykryta interpretacja semantyczna rozważanej sceny (pochodząca – przypomnijmy to – ze zbioru możliwych wniosków podanych przez eksperta na etapie akwizycji jego wiedzy w celu jej implementacji w systemie) może być traktowana jako poważna propozycja sposobu rozumienia i septycznej interpretacji aktualnie analizowanej sceny.

Podejście to będzie badane, doskonalone i konfrontowane z potrzebami praktyki, ale szkielet koncepcyjny tego modułu został tu zaprezentowany w całości.

Literatura

[1] Cpałka K.: *Zagadnienie interpretowalności wiedzy i dokładności działania systemów decyzyjnych*. EXIT, Warszawa 2009.

[2] Flasiński M.: *Wstęp do sztucznej inteligencji*. PWN, Warszawa 2011.

[3] Honglian Ma, Hunchuan Lu, Mingxiu Zhang: *A Real-time Effective System for Tracking Passing People Using a Single Camera*. Proceedings of the 7th World Congress on Intelligent Control and Automation, China 2008.

[4] Jankowski N.: *Meta-uczenie w inteligencji obliczeniowej*. EXIT, Warszawa 2011.

[5] Kisielewicz A.: *Sztuczna inteligencja i logika*. WNT, Warszawa 2011.

[6] Breitenstein M.: *Online Multiperson Tracking-by-Detection from a Single*. Uncalibrated Camera; IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 9, 2001, p. 1820–1833.

[7] Nowicki R.: *Rozmyte systemy decyzyjne w zadaniach z ograniczoną wiedzą*. EXIT, Warszawa 2009.

[8] PETS: Proceedings of Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, 2009.

[9] Regazzoni C.S., Cavallaro A., Wu Y., Konrad J., Hampapur A.: *Video Analytics for Surveillance: Theory and Practice*. Signal Processing Magazine, IEEE Volume: 27, 2010, p. 16–17.

[10] Rutkowski L.: *Metody i techniki sztucznej inteligencji*, PWN, Warszawa 2011.

[11] Stąpor K.: *Metody klasyfikacji obiektów w wizji komputerowej*. PWN, Warszawa 2011.

[12] Tadeusiewicz R., Ogiela M.R.: *Medical Image Understanding Technology, Series: Studies in Fuzziness and Soft Computing*, Vol. 156, Springer-Verlag, Berlin – Heidelberg – New York 2004.

[13] Tadeusiewicz R., Szczepaniak P.S.: *Basic Concepts of Knowledge-Based Image Understanding*. Chapter in book: Nguyen N.T., Jo G.S., Howlett R.J., Jain L.C. (eds.): *Agent and Multi-Agent Systems: Technologies and Applications, Lecture Notes on Artificial Intelligence*, vol. 4953, Springer-Verlag, Berlin – Heidelberg – New York 2008, pp. 42–52.

[14] Szczepaniak P.S., Tadeusiewicz R.: *The Role of Artificial Intelligence, Knowledge and Wisdom in Automatic Image Understanding*. Journal of Applied Computer Science, Vol. 18, No. 1, 2010, pp. 75–85.

[15] Tadeusiewicz R.: *Automatyczne rozumienie obrazów przez komputer jako element systemu e-kształcenia*. Rozdział w pracy zbiorowej Migdałek J., Folta W. (red.): *Technologie informacyjne w warsztacie nauczyciela*. Księgarnia Akademicka, Kraków 2010, pp. 13–27.

[16] Tadeusiewicz R., Mikrut Z.: *Wymogi czasu rzeczywistego w systemach wizyjnych specjalnego przeznaczenia*. Rozdział nr 44 w pracy zbiorowej: Trybus L., Samolej S. (red.): *Projektowanie, analiza i implementacja systemów czasu rzeczywistego*. WKiŁ Warszawa 2011, pp. 525–539.

prof. zw. dr hab. inż. Ryszard Tadeusiewicz
Akademia Górniczo-Hutnicza
im. Stanisława Staszica w Krakowie;
Wydział Elektrotechniki, Automatyki,
Informatyki i Inżynierii Biomedycznej;
Katedra Automatyki i Inżynierii
Biomedycznej