

Politechnika Poznańska
Wydział Informatyki
Katedra Sterowania i Inżynierii Systemów
Pracownia Układów Elektronicznych i Przetwarzania Sygnałów

Interfejsy człowiek-komputer do wspomagania automatycznego rozpoznawania zagrożeń

Streszczenie rozprawy doktorskiej

Julian Balcerek

Promotor: prof. dr hab. inż. Adam Dąbrowski
Promotor pomocniczy: dr inż. Paweł Pawłowski

Poznań 2016

1. Wprowadzenie

We współczesnym świecie, a szczególnie w środowiskach zurbanizowanych, gdzie występuje bardzo duże zagęszczenie ludności, obserwuje się zwiększoną liczbę zagrożeń i związanych z nimi zdarzeń, w tym kolizji i wypadków komunikacyjnych, wykroczeń, występków i zbrodni [EurC2011, EurC2014, Stat2015].

Standardowe systemy przeciwdziałania zagrożeniom w środowiskach zurbanizowanych działają następująco: po wystąpieniu zdarzenia są zbierane informacje o zagrożeniu, np. w postaci danych audiowizualnych lub tekstowych. W prostych sytuacjach coraz częściej stosowane są algorytmy automatyczne, np. wykrywany jest ruch, który uruchamia alarm. Sytuacje bardziej złożone jak np. zdarzenie drogowe wymagają interwencji odpowiednio przygotowanego operatora, najczęściej centrum ratunkowego. Operator taki posiada wiedzę ekspercką, analizuje i wybiera odpowiednie informacje, a następnie podejmuje ostateczną decyzję, np. o wysłaniu karetki pogotowia. Adekwatna reakcja prowadzi w efekcie do eliminacji zagrożenia.

W niniejszej rozprawie doktorskiej autor skoncentrował się na dwóch najważniejszych systemach przeciwdziałania zagrożeniom: systemie monitoringu wizyjnego i systemie telefonicznego numeru alarmowego.

W systemach monitoringu wizyjnego obserwuje się szybki i intensywny rozwój infrastruktury i algorytmów przetwarzania danych [McCa2002, Pozn2016]. W skład infrastruktury systemu monitoringu wchodzi nie tylko kamery, ale także urządzenia rejestrujące, stacje bazowe i węzłowe. Informacje są zbierane statycznie (w przypadku kamer stacjonarnych) lub dla zdefiniowanego punktu widzenia (w przypadku kamer obrotowych). Ze względu na ogromną ilość danych powstaje duże zapotrzebowanie na nowe techniki wizualizacji i analizy danych multimedialnych [Velt2012]. W związku z tym, rozwój nowych narzędzi wspomagających wykrywanie zagrożeń, wykorzystujących system wizyjny, staje się ważnym zadaniem i istotnym problemem badawczym.

Obecnie dobrze rozwinięte są tzw. klasyczne metody detekcji zagrożeń w systemach wizyjnych. Są to m.in.: detekcja ruchu, wykrywanie pozostawionych przedmiotów, dymu, ognia i przechodzenia na czerwonym świetle, śledzenie i zliczanie obiektów, np. osób lub pojazdów, rozpoznawanie różnych parametrów biometrycznych, w tym twarzy, płci, analiza chodu, a także rozpoznawanie numerów tablic rejestracyjnych, wykrywanie zatorów drogowych, zabronionych manewrów, wypadków drogowych, itp. [Data2012, Cetn2012, Dabr2010, Chen2012, Zhao2012].

Zostało dowiedzione, że klasyczne metody automatycznej detekcji zagrożeń przez systemy wizyjne poprawiają wydajność operatora dla średnio skomplikowanych zadań. Jednak operator skupiony na prostym zadaniu ma większą wydajność nawet bez wspomaganie [Rank2012]. Ponadto, obszary rozpraszające uwagę można ograniczać przybliżając jedynie prostokątne fragmenty obrazu, a nie precyzyjnie wybrane obiekty sceny. Aby wielogodzinna, monotonna i nużąca obserwacja obrazu przez operatora monitoringu wizyjnego mogła stać się bardziej atrakcyjna, a zarazem dokładniejsza, należałoby zastosować niestandardowe techniki wizualizacji takie jak np. stereowizja.

Obecnie stereowizję stosuje się w technice do wizualizacji dla człowieka oraz do akwizycji i przetwarzania danych przez maszyny. Informacje są pozyskiwane na podstawie dwóch kanałów obrazu i różnic pomiędzy nimi. W dziedzinie automatyki i robotyki systemy

stereowizyjne są używane do nawigowania, wykrywania obiektów i nawierzchni drogi [LinC2005, Bota2009, Wang2013].

Współczesna interaktywna telewizja 3D, będąc przykładem interfejsu człowiek-komputer, powoduje skupienie uwagi, poczucie obecności i zagłębienie widza w scenie [Huyn2011]. W przypadku operatora monitoringu zastosowanie takiej wizualizacji 3D pozwoliłoby na zachowanie skupienia i koncentrację na obserwowanej sytuacji i tym samym dokładniejszą analizę. Dzięki stereowizji percepcja głębi przez człowieka jest znacznie lepsza niż w przypadku obrazów 2D i jest on w stanie lepiej określić rozmiar obiektu w 3D niż w 2D [Huyn2011, LuoK2007]. Zaistniała potrzeba przeprowadzenia eksperymentów rozpoznawania szczegółów sytuacji przez operatora monitoringu takich jak liczenie osób lub szacowanie odległości, co może mieć istotny wpływ na decyzje podejmowane w sytuacjach zagrożenia. Aby operator nie był nadmiernie obciążony, wizualizacja 3D powinna być opcją w sytuacjach, które wymagają szczególnej uwagi [Huyn2011].

Efekt 3D może być wizualizowany dla człowieka przy pomocy różnych technik. Wśród nich są ekrany autostereoskopowe nie wymagające używania okularów [Urey2011]. W przypadku wizualizacji stereowizyjnej można stosować konwersję obrazu monoskopowego na stereoskopowy. Komercyjne metody stereowizyjne, stosowane w celach rozrywkowych, stosują konwersję zorientowaną na efekty i znacznie obciążają układy graficzne [Bart2008]. Tak więc, wydajne metody konwersji 2D do 3D są ważnym zagadnieniem i zaistniała potrzeba przygotowania prostego i efektywnego narzędzia do konwersji obrazów ze sterowaniem postrzeganą głębią [TamZ2006, KimD2007, LiWa2010, Chan2012]. Technika stereowizyjna dla operatora monitoringu może towarzyszyć i jest komplementarna z innymi metodami przetwarzania obrazów, takimi jak usuwanie szumu z obrazów cyfrowych lub stosowanie obrazowania o dużym zakresie jasności dla zdjęć uzyskanych w trudnych warunkach oświetleniowych [Koni2014, Koni2016].

Niestety, ze względu na ograniczenia w postaci wysokiej ceny sprzętu, niedostatecznej separacji kanałów obrazu dla większych odległości i trudności przy stosowaniu efektów specjalnych, wciąż duża liczba zdjęć i materiałów wideo jest przygotowywana w technice monoskopowej.

Drugi sposób reagowania i eliminacji zagrożeń jest oparty na zgłoszeniach osób w systemach telefonicznego numeru alarmowego. Rozwinięty w Polsce system powiadamiania ratunkowego wykorzystuje numer 112, który jest standardem w krajach Unii Europejskiej [Supr2013, OBri2013].

Wiele problemów technicznych i organizacyjnych jest związanych z usługami telefonicznych numerów alarmowych i systemami pokrewnymi. Na przykład, odpowiednia reakcja musi być oparta na prawdziwych, istotnych i spójnych danych. Niestety, w praktyce jedno zdarzenie może być zgłaszane wielokrotnie, a osoby zgłaszające mogą podawać nieprecyzyjne lub nieprawidłowe dane. Dodatkowo, operator może popełniać błędy podczas wprowadzania danych i ich analizy. Ponadto, usługi telefonu alarmowego są często nadużywane – większość zgłoszeń jest niezasadnych, a część nawet fałszywych [ITPC2012].

Infrastruktura techniczna i organizacyjna służb ratunkowych jest wciąż rozwijana i udoskonalana. Dotychczas połączona została obsługa zgłoszeń i organizacja działalności ratowniczej [Ziaj2011]. Wiele technicznych możliwości wspierania operatorów zastosowano już w dostępnych na rynku systemach [Buch2013]. Przykładowo, operator ma możliwość identyfikacji numeru telefonu oraz ma m.in. dostęp do listy jednostek ratowniczych, katalogu

substancji toksycznych, neutralizujących i środków gaśniczych. Podstawowe informacje o każdym zgłoszeniu są zapisywane w formie tekstowej i głosowej. Niestety, informacje o lokalizacji mogą być niedokładne, a zapisana historia niewiarygodna ze względu na możliwość wystąpienia fałszywych danych [ECCC2015]. Z drugiej strony, manualne przeszukiwanie danych jest czasochłonne i kosztowne.

Operator Centrum Powiadamiania Ratunkowego (CPR) ma możliwość reagowania na zagrożenia. Jest jednak problem z oceną zasadności reakcji i poprawnym doborem jednostek ratowniczych. Ważnymi zagadnieniami jest rozpoznawanie zgłoszeń dotyczących tej samej sytuacji, a raportowanych wcześniej przez innych dzwoniących oraz weryfikacja, czy już wcześniej dane zdarzenie było zgłaszane. Istotnymi zagadnieniami są również: rozpoznawanie czy dany dzwoniący kiedykolwiek wcześniej dzwonił na numer alarmowy i wyszukiwanie powtarzających się nieuzasadnionych zgłoszeń tych samych osób.

Mimo, że systemy telefonicznego numeru alarmowego są powszechnie używane, szczegóły zastosowanych algorytmów przetwarzania nie są publicznie dostępne. Wśród nielicznych opisanych w literaturze przedmiotu systemów, wykrywane były zgłoszenia dotyczące huraganu na podstawie trzech oryginalnych cech zdarzenia [Klem2009] lub rozmówcy byli identyfikowani na podstawie cech głosu [Grzy2014]. Dokładne rozpoznanie mówcy lub zdarzenia wymaga wyboru, przetwarzania i porównania istotnych informacji uzyskanych z dostępnych źródeł. Opis danych w celu rozpoznawania powinien zawierać precyzyjnie określone metadane [Reve2008, Rawa2010, Sici2014]. Spośród wszystkich reprezentacji powinny zostać wybrane te najlepiej modelujące rozważane obiekty [Krie2006, Rash2009].

Głównymi źródłami informacji w systemach telefonicznego numeru alarmowego są: raport operatora centrum alarmowego, rejestrator rozmów, usługi operatora telekomunikacyjnego, transkrypcje i wyniki wielopoziomowej analizy dźwięku [Dabr2012, Drga2015]. Typowo, podczas rozmowy telefonicznej operator zadaje serię pytań na temat przebiegu, miejsca i czasu zdarzenia oraz danych rozmówcy. Na podstawie rozmowy można również uzyskać informacje na temat tła akustycznego oraz płci, wieku i szybkości mówienia rozmówcy jak innych cech wypowiedzi takich jak występowanie błędów lingwistycznych, obcego akcentu, jąkania, powtarzania słów, chryпки, pauz wypełnionych i błędów logicznych.

Rozpoznawanie najbardziej podobnych danych może być rozumiane jako poszukiwanie obiektów najbliższych obiektowi odniesienia, przy czym odległości mogą być rozważane dla klas binarnych, numerycznych i opartych na terminach [Alla2007, Gant1999]. Można również stosować punktację, gdzie najbardziej podobne są obiekty o najwyższej punktacji [Berr2003]. Istotność cech z punktu widzenia wyszukiwania powinna być odzwierciedlona w ich wagach [Krie2006].

Porównywanie rozmów nie jest zadaniem trywialnym, ponieważ, nawet jeżeli cechy mają różne wartości, nie oznacza to, że odnoszą się do różnych osób. Przykładowo: lokalizacja zdarzenia lub numer telefonu, a także grupa wiekowa, do której należy rozmówca mogą być zmienne. Z drugiej strony, zgłoszenia o wielu tych samych wartościach cech, niekoniecznie muszą dotyczyć tego samego zdarzenia.

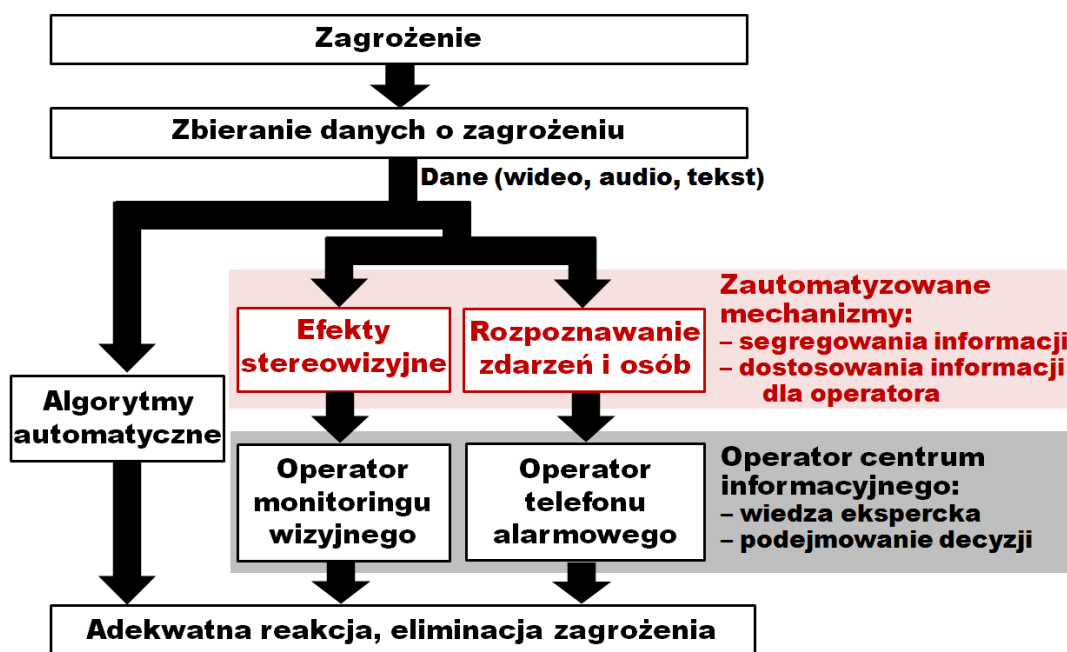
Interfejsy człowiek-komputer, które autor rozwijał w pracy, wspomagają automatyczne rozpoznawanie zagrożeń. Skoncentrowano się na zagrożeniach występujących w środowiskach zurbanizowanych. W badaniach podjęto konkretne problemy występujące w Poznaniu i w Wielkopolsce. Występują one jednak i są istotne w całym współczesnym świecie.

2. Cele, teza naukowa i organizacja pracy

Zostały sformułowane dwa cele pracy. Poznawczym celem badań było opracowanie zautomatyzowanych mechanizmów przetwarzania informacji wizyjnej i audio do wspomagania operatorów centrów informacyjnych. Celem aplikacyjnym było opracowanie oprogramowania do uzyskiwania efektu stereowizyjnego w monitoringu miejskim oraz rozpoznawania zdarzeń i osób w systemach telefonów alarmowych.

Teza naukowa rozprawy doktorskiej została sformułowana następująco: opracowane interfejsy człowiek-komputer (stereowizja w monitoringu wizyjnym, rozpoznawanie zdarzeń i osób na podstawie rozmów telefonicznych na numery alarmowe) wspomagają pracę operatorów centrów informacyjnych i poprawiają bezpieczeństwo w obszarach zurbanizowanych.

Autor, jako koncepcję rozwiązania problemu badawczego, zaproponował modyfikację systemów przeciwdziałania zagrożeniom w celu wspomagania operatorów centrów informacyjnych, którzy są znacznie obciążeni pracą, trudnościami i odpowiedzialnością (rys. 1). W rozprawie autor wprowadził wizualizację stereowizyjną dla operatora monitoringu wizyjnego (w tzw. podsystemie wizyjnym) oraz mechanizm rozpoznawania zdarzeń i osób dla operatora telefonicznego numeru alarmowego (w tzw. podsystemie telefonicznego numeru alarmowego).



Rys. 1. Systemy przeciwdziałania zagrożeniom w środowisku zurbanizowanym

Autor udowodnił tezę przez zaproponowanie schematów konwersji obrazów 2D na 3D i mechanizmów rozpoznawania i wyszukiwania zdarzeń i mówców oraz liczne eksperymenty przy użyciu zarejestrowanych obrazów z monitoringu wizyjnego i rozmów na telefon alarmowy.

Praca została zorganizowana w następujący sposób. W pierwszym rozdziale został opisany obszar badań, cele, teza naukowa i zakres badań. Problemy, które zostały rozwiązane w pracy, odniesiono do zastanego stanu wiedzy. Następnie autor przedstawił swoje główne

osiągnięcia naukowe. W drugim rozdziale przedstawiono mechanizmy percepcji, generowania i wizualizacji wrażeń stereowizyjnych dla człowieka. Autor zaproponował schematy wywoływania wrażeń stereowizyjnych na podstawie obrazów monoskopowych i rozwiązania problemu wypełniania luk informacyjnych w wygenerowanych obrazach 3D. W tym samym rozdziale opisano również eksperymenty dotyczące postrzegania efektu 3D przez człowieka [Balc2012, Balc2014a] i zaproponowano podejście do szybkiej konwersji obrazów 2D do 3D opartej na zredukowanej liczbie parametrów kontrolnych. Konwersja ta jest dedykowana operatorowi monitoringu. Dodatkowy eksperyment dotyczący jakości i głębi obrazów zweryfikował zaproponowane schematy wywoływania wrażeń stereowizyjnych [Balc2014]. Rozdział trzeci zawiera porównanie wyników rozpoznawania ważnych szczegółów sytuacji w scenach z monitoringu przez wytrenowanych i niewytrenowanych obserwatorów. W rozdziale czwartym autor zaproponował mechanizm automatycznej selekcji informacji z rozmów na telefony alarmowe. W celu zwiększenia jakości rozpoznawania, procedury zostały rozszerzone na wielowymiarowe zależności pomiędzy cechami. Ponadto, autor przedstawił procedurę uczenia maszynowego dla wytrenowania parametrów procedury wyszukiwania i automatyczną klasyfikację rozmów przez sztuczne sieci neuronowe. Rozdział piąty został poświęcony eksperymentom rozpoznawania osób dzwoniących i zdarzeń przy użyciu zaproponowanego mechanizmu wyszukiwania. Przygotowano zbiór zgłoszeń na numer alarmowy wraz z ich opisem. Zestawy parametrów mechanizmu wyszukiwania zostały wybrane i wytrenowane przy użyciu procedury uczącej. Wykonano studium przypadków automatycznego rozpoznawania nieprawidłowych przypadków przy pomocy zależności pomiędzy cechami [Balc2015, Balc2015a]. Przeprowadzono również serię eksperymentów klasyfikacji rekordów rozmów przy użyciu sieci neuronowych. W ostatnim rozdziale autor przedstawił wnioski i podsumował przeprowadzone badania.

3. Główne osiągnięcia naukowe

Główne osiągnięcia naukowe autora obejmują szereg innowacyjnych modyfikacji dwóch najważniejszych systemów zapobiegania zagrożeniom, czyli monitoringu dozorowego (CCTV, ang. closed-circuit television) oraz telefonu alarmowego, w odniesieniu do właściwego przedstawiania tych informacji dla operatora. Obejmują one wizualizację stereowizyjną dla operatora monitoringu wizyjnego oraz mechanizm rozpoznawania zdarzeń i osób dla operatora systemu telefonicznego numeru alarmowego.

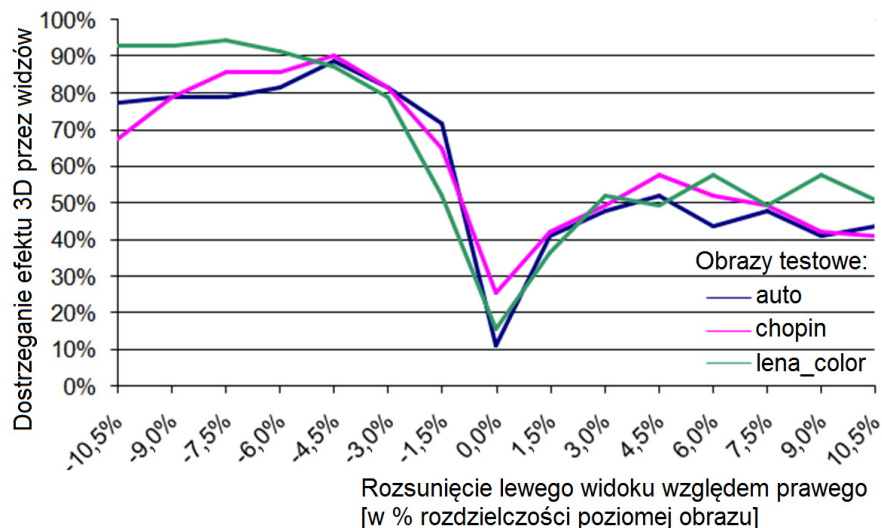
3.1. Wywoływanie wrażeń stereowizyjnych

W ramach podsystemu wizyjnego przebadane zostały mechanizmy wywoływania, procedury generowania i postrzeganie przez człowieka efektów 3D. Badania dotyczyły subiektywnego odbioru obrazów stereowizyjnych przez widzów na ekranie.

Dobrze znane komercyjne metody stereowizyjne stosowane np. w telewizji nie mogą być bezpośrednio wykorzystywane w monitoringu, ponieważ są one za bardzo zorientowane na wywoływanie widowiskowych efektów 3D. Metody stereowizyjne dla systemów monitoringu powinny oferować efekty 3D, które są wiarygodne i nie zakłócają odbioru istotnych szczegółów obrazu. Co więcej, ich intensywność powinna być całkowicie pod kontrolą

operatora. Tak więc, badania i eksperymenty zostały skoncentrowane głównie na badaniu postrzegania efektu stereowizyjnego przez człowieka.

Obraz stereowizyjny powstaje na podstawie dwóch widoków (obrazów monoskopowych): dla lewego i prawego oka. Eksperymentalnie udowodniono, że przedstawienie jednego obrazu monoskopowego dla prawego oka, a jego przesuniętej w poziomie kopii dla lewego oka, wyzwała w mózgu wrażenie 3D. Badania na grupie 71 widzów zaprezentowano na rys. 2. Przy odpowiednich wartościach przesunięcia (4,5% rozdzielczości poziomej obrazu w lewo) ok. 90% widzów dostrzegało efekt stereowizyjny [Balc2012].



Rys. 2. Dostrzeganie efektu 3D przez widzów w zależności od rozsunięcia widoków obrazu

Bardziej naturalne wrażenia 3D można uzyskać dla poziomych przesunięć obiektów występujących w obrazie. Nawet najprostszy przypadek mapy głębi, tj. mapa binarna, jest wystarczający do wywołania efektu 3D. Taki efekt 3D, wskazując na obiekt zainteresowania, może zwrócić szczególną uwagę operatora monitoringu. Zaproponowano generowanie wrażeń 3D ze źródłowych obrazów 2D wraz ze schematami konwersji 2D do 3D [Balc2008, Balc2011b, Balc2012]. Zaproponowano także rozwiązanie problemu tzw. luk informacyjnych przy krawędziach obiektów, czyli obszarów sceny w wygenerowanym wirtualnym widoku, które nie są dostępne z perspektywy rzeczywistego widoku, a pogarszają percepcję obrazu 3D. Mogą one być wypełniane metodą interpolacyjną lub odbicia lustrzanego i filtracji obrazu [Balc2008, Balc2011b].

W rozpatrywanym podsystemie monitoringu wizyjnego zaproponowane użycie efektu stereowizyjnego polega na przybliżaniu wybranych elementów sceny w kierunku operatora w zrównoważonym efekcie 3D. W tym celu zaimplementowano i przetestowano szybkie i skuteczne (tj. operujące w czasie rzeczywistym) algorytmy do konwersji obrazów 2D do 3D, oparte na uproszczonych mapach głębi [Balc2014a]. Pomimo, że metody przetwarzania obrazu zastosowane w tej konwersji oferują jedynie przybliżony do naturalnego efekt 3D, biorąc pod uwagę zdolności ludzkiej percepcji wzrokowej, jest on wystarczający.

Kolejne testy miały na celu zaproponowanie najlepszych parametrów sterujących procedurami generowania efektów 3D [Dabr2012b, Balc2014a]. Korzystając z regresji liniowej, na podstawie testów przeprowadzonych na 56 widzach, została ustalona liniowa

zależność pomiędzy parametrami sterującymi. Na rys. 3 zostały przedstawione zależności pomiędzy parametrami sterującymi dla schematu konwersji opartego na bezpośrednim przesunięciu obszarów wybranego obrazu testowego z lukami informacyjnymi wypełnionymi za pomocą interpolacji.



Rys. 3. Rozkład najlepszych wartości parametrów przesunięcia obiektu względem przesunięcia tła dla schematu bezpośredniego z interpolacją dla wybranego obrazu testowego

Interesujące jest, że po przybliżeniu omawianej zależności prostą, ponad 90% spośród 25 badanych potrafiło wskazać jeden punkt na tej prostej oznaczający najlepiej postrzegany efekt 3D. Tak więc, jest możliwe wywoływanie wiarygodnych efektów 3D za pomocą uproszczonych procedur, a zależności funkcyjne między parametrami sterującymi efektem 3D pozwalają na zmniejszenie liczby tych parametrów [Dabr2012b].

Autor przeprowadził również serię eksperymentów dotyczących jakości i intensywności efektu 3D używając zdjęć podobnych do typowo analizowanych w systemach CCTV [Balc2013, Balc2014]. Wyniki eksperymentów przeprowadzonych na dwóch trzydziestoosobowych grupach widzów, są przedstawione w tabeli 1.

Tabela 1. Wyniki eksperymentu dotyczącego jakości i intensywności efektu 3D (w skali MOS)

Wariant obrazu	Obrazy 2D	Obrazy 3D wygenerowane z obrazów 2D	Rzeczywiste obrazy 3D	Obrazy 3D ze zwiększoną głębią
Jakość całego obrazu	4,5	3,5	4,5	3,5
Intensywność efektu 3D	1,4	3,3	2,8	3,6

Widać, że zaproponowane schematy konwersji zwiększają zakres głębi zachowując akceptowalną jakość obrazu [Balc2013, Balc2014]. Dobrze widoczna głębia w sztucznie wygenerowanych obrazach 3D (z obrazów 2D) jest lepiej dostrzegalna niż głębia rzeczywistych obrazów 3D. Schematy dodatkowego zwiększania głębi zastosowane na rzeczywistych obrazach 3D dają bardziej intensywne wrażenie głębi niż źródłowe obrazy 3D.

Po zastosowaniu przekształceń postrzegana odległość od obserwatora do obserwowanych obiektów, zgodnie z założeniami, zmniejszyła się względem tła kosztem nieznacznie zmniejszonej jakości obrazu. Należy jednak zaznaczyć, że analizowane obiekty nie były zniekształcone, co jest ważną zaletą w systemach CCTV.

3.2. Rozpoznawanie zdarzeń w trójwymiarowym monitoringu wizyjnym

Autor spodziewał się, że dzięki wizualizacji 3D opartej na dwóch kanałach obrazu, szczegółowa inspekcja wizyjna, polegająca np. na liczeniu osób lub szacowaniu odległości pomiędzy obiektami, powinna być szybsza i bardziej dokładna niż w przypadku obrazów 2D. Te zalety wizualizacji stereowizyjnej zostały w pełni potwierdzone w przeprowadzonych eksperymentach.

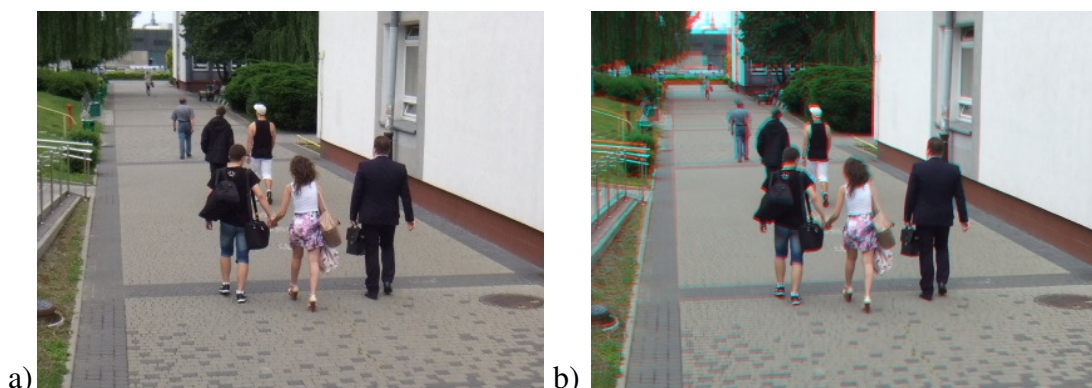
Przebadano rozpoznawanie istotnych szczegółów w wybranych scenach 3D z systemu CCTV w porównaniu do analogicznych scen 2D. W serii eksperymentów brało udział 48 niewytrenowanych i 6 wytrenowanych obserwatorów. Przygotowano w sumie 24 zdjęcia testowe, każde w wersji 2D i 3D.

Potwierdzono, że postrzeganie detali obiektu w obrazie po zastosowaniu zaproponowanego schematu konwersji 2D do 3D jest w pełni zachowane. Wprowadzony efekt 3D uwydatnia wybrany obiekt będący przedmiotem zainteresowania, co może być traktowane jako wyróżniający marker.

Wykazano także, że stereowizyjne pozyskiwanie obrazów i ich wizualizacja pozwalają dostrzec istotne szczegóły obrazów 3D, które nie są dokładnie widoczne w obrazach 2D.

Również oszacowanie odległości między obiektami, w tym pomiędzy ludźmi, podczas obserwacji stereoskopowej jest, w większości przypadków, o wiele lepsze niż w trybie monoskopowym. Stereowizja może również pomóc w oszacowaniu trajektorii i relacji pomiędzy obiektami, które mogą stanowić potencjalne zagrożenie.

Przykładowy obraz testowy w wersji 2D i 3D wykorzystany w eksperymencie dotyczącym liczenia osób, przedstawiono na rys. 4.



Rys. 4. Przykładowy obraz testowy dla eksperymentu liczenia osób w wersji a) 2D i b) 3D

Dzięki stereowizji można poprawić dokładność liczenia osób, ale zajmuje ono nieznacznie więcej czasu w przypadku niewytrenowanych widzów. Z drugiej strony, w przypadku wytrenowanych widzów, wyniki są lepsze w obu aspektach: dokładności i szybkości analizy. Wyniki wybranych eksperymentów zebrano w tabeli 2.

Tabela 2. Wybrane wyniki eksperymentów dotyczących dokładności i szybkości manualnej analizy typowych sytuacji w monitoringu wizyjnym

Typ analizy	Obrazy 2D		Obrazy 3D	
Dostrzeganie istotnych szczegółów niewidocznych w 2D	Poprawne odpowiedzi: 0%		Poprawne odpowiedzi: 93,8%	
Szacowanie odległości pomiędzy obiektami	Błąd względny: 29,5%		Błąd względny: 20,1%	
Liczenie osób (obserwatorzy niewytrenowani)	Czas: 8,2 s	Poprawne odpowiedzi: 91,7%	Czas: 10,8 s	Poprawne odpowiedzi: 100%
Liczenie osób (obserwatorzy wytrenowani)	Czas: 4,4 s	Poprawne odpowiedzi: 33,3%	Czas: 4,1 s	Poprawne odpowiedzi: 54,8%

Podsumowując, wizualizacja stereowizyjna poprawia rozpoznawanie obserwowanych szczegółów w obrazie. Autor doświadczalnie udowodnił, że zautomatyzowane obrazowanie 3D poprawia dokładność i szybkość manualnej analizy typowych sytuacji w systemie CCTV. Zatem proponowane obrazowanie 3D może być oferowane jako cenna opcja w systemach monitoringu.

3.3. Mechanizmy selekcji i porównywania metadanych pozyskanych z rozmów na telefoniczny numer alarmowy

W drugim podsystemie przeciwdziałania zagrożeniom, związanym z usługami telefonów alarmowych, autor zaproponował mechanizm rozpoznawania osób i zdarzeń (por. rys. 1). Rozpoznawanie zagrożeń w systemie telefonicznego numeru awaryjnego obejmuje rozpoznawanie zdarzeń zgłoszonych wcześniej przez konkretnego rozmówcę, a także sprawdzenie, czy zdarzenie nie zostało już wcześniej zgłoszone.

Osiągnięcie precyzyjnych wyników rozpoznawania mówcy lub zdarzenia wymaga wyboru oraz porównania istotnych i wiarygodnych cech rozmów pozyskiwanych z dostępnych źródeł [Balc2009]. Zaproponowany mechanizm operuje na metadanych pobranych ze starannie dobranych 35 cech rozmów, rozmówców, bądź zdarzeń. Cechy te są związane m.in. z lokalizacją, czasem, szybkością mówienia, wiekiem dzwoniącego, występowaniem błędów językowych, powtarzaniem słów i innymi właściwościami charakterystycznymi dla mówcy lub zdarzenia. Rozpatrywane cechy mogą być reprezentowane za pomocą binarnych, numerycznych lub tekstowych typów danych.

Zaproponowany mechanizm rozpoznawania jest oparty na sekwencjach porównań wartości cech bieżącego rekordu zgłoszenia (rekordu odniesienia) z wartościami cech innych rekordów zgłoszeń, zapisanych i przechowywanych w bazie danych. Zestaw zaproponowanych procedur zawiera porównania binarne, numeryczne i oparte na nazwach, co w połączeniu z pozostałymi mechanizmami zapewnia dużą niezawodność systemu [Balc2011a]. W tabeli 3 przedstawiono typy porównań metadanych, odpowiednie wzory i przykłady ich zastosowań.

Tabela 3. Typy porównań metadanych, odpowiednie wzory i przykłady ich zastosowań

Rodzaj porównania metadanych	Wzór	Przykłady porównywanych cech
Binarne	$v_B(f_j) = \begin{cases} 1 \text{ for } a_j = b_{ij} \\ 0 \text{ for } a_j \neq b_{ij} \end{cases}$	Miasto (binarna zgodność lub niezgodność), chrypka, obcy akcent
Numeryczne	$v_N(f_j) = \begin{cases} v_1 \text{ for } \langle 0, d_1 \rangle \\ \dots \\ v_k \text{ for } \langle d_{k-1}, d_k \rangle \\ \dots \\ v_R \text{ for } \langle d_{R-1}, d_{\max} \rangle \end{cases}$ <p>gdzie $d = a_j - b_{ij} \in \langle d_{\min}, d_{\max} \rangle$</p>	Data, godzina i czas trwania zdarzenia, długość rozmowy
Oparte na nazwach	$P_j = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{bmatrix}$ $v_T(f_j) = p_{mn}$	Miasto (uwzględnione czas i odległości drogowe), wiek dzwoniącego (reprezentowany przez przedział wiekowy), szybkość mówienia
Oparte na korelacjach	$P_j(f_k) = \begin{bmatrix} p_{11}(f_k) & p_{12}(f_k) & \dots & p_{1N}(f_k) \\ p_{21}(f_k) & p_{22}(f_k) & \dots & p_{2N}(f_k) \\ \dots & \dots & \dots & \dots \\ p_{N1}(f_k) & p_{N2}(f_k) & \dots & p_{NN}(f_k) \end{bmatrix}$ $v_C(f_j, f_k) = p_{mn}(f_k)$	Zmiana wieku zgłaszającego w czasie, zmiana lokalizacji w czasie
Punktacja globalna	$S_i = \sum_{f_j \in F_B} w_f v_B(f_j) + \sum_{f_j \in F_N} w_f v_N(f_j) + \sum_{f_j \in F_T} w_f v_T(f_j) + \sum_{f_j, f_k \in F_C, f_j \neq f_k} w_{f_j f_k} v_C(f_j, f_k)$	Wszystkie cechy

Zakładając, że a_j jest parametrem j -tej cechy f_j bieżącego zgłoszenia, a b_{ij} jest odpowiadającym parametrem tej samej cechy dla i -tej konwersacji w bazie, dla każdej pary a_j i b_{ij} można określić wartości podobieństw. Odpowiadają one prawdopodobieństwu tego, że analizowane rozmowy odnoszą się do tego samego zdarzenia lub tego samego zgłaszającego [Balc2015].

Zdefiniowano dodatkowe zależności pomiędzy porównaniami dwóch cech f_j i f_k , zwane dalej korelacjami, które pomogły zwiększyć wydajność systemu [Balc2010]. Odpowiednie wzory również przedstawiono w tabeli 3. Przykładem tego typu porównania może być problem rozpoznawania mówców na podstawie rozmów zgłoszonych z różnych miast, kiedy mamy do czynienia z procesem dynamicznym. Wraz z upływem czasu rośnie prawdopodobieństwo, że dwie rozmowy zostały wykonane przez tę samą osobę, gdyż miała ona coraz więcej czasu na przemieszczenie się pomiędzy miastami. Korelacja lokalizacji z czasem uwzględnia taką zależność i decyduje o przydzieleniu adekwatnego

prawdopodobieństwa. Dzięki temu można określić, że następujące po sobie w krótkim czasie zgłoszenia dotyczące odległych zdarzeń mogą być fałszywe i dany mówca nie podaje wiarygodnych informacji.

Innym przykładem korelacji jest uwzględnienie upływu czasu pomiędzy zgłoszeniami i aktualnej grupy wiekowej, do której należy rozmówca. Na przestrzeni lat człowiek zmienia swój wiek i pomimo, że w dwóch nagraniach została zarejestrowana osoba w różnym wieku to korelacja wieku z czasem pozwala określić prawdopodobieństwo, że jest to ta sama osoba.

W procesie rozpoznawania, globalna punktacja, wyznaczona jako suma ważonych częściowych punktacji dopasowania, jest obliczana dla każdej porównywanej konwersacji i przedstawiana operatorowi z określonym procentem dopasowania do rozmowy odniesienia. Wartości podobieństw i wag mogą być ustalane na podstawie analizy statystycznej lub metod uczenia maszynowego [Balc2015]. To dopasowanie procentowe określa pozycję rekordu rozmowy na liście rankingowej wyników. Najlepsze wyniki rozpoznawania są dla rekordów z najwyższymi wartościami punktacji dopasowania (por. rys. 5).

Dodatkowe rozszerzenie, jakim jest graficzna interpretacja różnych rodzajów procedur wyszukiwania, podobna do technik przetwarzania obrazu, zostało przedstawione w [Balc2015] i [Balc2015a].

Wprowadzenie zależności między więcej niż dwoma różnymi cechami oferuje znacznie bardziej precyzyjne modelowanie rzeczywistych, skomplikowanych relacji. W tym celu, autor rozszerzył mechanizm rozpoznawania oparty na korelacjach pomiędzy dwoma cechami na wielowymiarowe porównania dotyczące dowolnej liczby cech. Odpowiednie wzory zostały uogólnione i rozszerzone na wielowymiarową przestrzeń cech. Takie podejście zapewnia bardziej szczegółowy opis zdarzeń i poprawia wyniki rozpoznawania. Ponadto, pozwala skalować precyzję rozpoznawania w zależności od danego przypadku i dostępnej szczegółowości opisów w bazie zgłoszeń alarmowych.

3.4. Rozpoznawanie osób i zdarzeń w systemie telefonicznego numeru alarmowego

W celu przetestowania zaproponowanego mechanizmu rozpoznawania, autor we współpracy z pracownikami i studentami Politechniki Poznańskiej przygotował eksperymentalny zbiór rozmów na telefon alarmowy. Ze względu na ograniczenia prawne, niemożliwe było przeprowadzenie badań na bazach danych rzeczywistych rozmów. W związku z tym należało samodzielnie jak najdokładniej odwzorować próbkę prawdziwej bazy połączeń alarmowych [Dabr2012]. W tym celu zbudowano system symulujący fragment CPR dotyczący telefonu alarmowego. W przeciągu siedmiu lat, 192 różne osoby zgłosiły 101 zdarzeń, a zgłoszenia obsłużyło czterech operatorów. W bazie [Balc2015a] zgromadzono 669 rozmów (pliki dźwiękowe i metadane) o łącznej długości ponad 21 godzin.

Przed wykonaniem zgłoszenia, rozmówca oglądał losowo wybrany film pochodzący z zestawu wcześniej przygotowanych krótkich filmów zawierających spreparowane sceny kryminalne. Zadaniem mówcy było zgłoszenie zaobserwowanej sytuacji. Rozmowy, jak w prawdziwej bazie danych CPR, różniły się długością, kompletnością, jakością i wiarygodnością opisu. W bazie znalazło się 23415 wartości cech, co oznaczało, że nie wypełniono 23,6% z możliwych wartości cech. Wartości tych nie podali rozmówcy lub nie

uzupełnili ich operatorzy. Zbiór rozmów naśladował prawdziwą bazę CPR także pod względem statystyk zdarzeń, tj. udziału interwencji, zdarzeń w ruchu drogowym i przestępstw wraz z ich podkategoriami w zgłoszeniach do CPR w Poznaniu w roku 2015.

Dodatkowo, bazę podzielono na dwa podzbiory uwzględniające różne konfiguracje wyszukiwania. W pierwszym podzbiorze 309 zgłoszeń dotyczyło jednego dnia w Poznaniu i odnosiło się do 51 zdarzeń w ośmiu lokalizacjach. W drugim podzbiorze występowało 50 zdarzeń, 18 różnych dat, i 360 zgłoszeń z 50 lokalizacji w Wielkopolsce, w tym osiem w Poznaniu i 42 w innych miastach.

Początkowe eksperymenty przeprowadzono z użyciem manualnie dobranych zestawów parametrów mechanizmu rozpoznawania. Dla najlepszego zestawu parametrów rozmówca został prawidłowo rozpoznany w co najmniej 78% przypadków, natomiast właściwe zdarzenie zostało znalezione w 89% przypadków (rys. 5a i 5c).

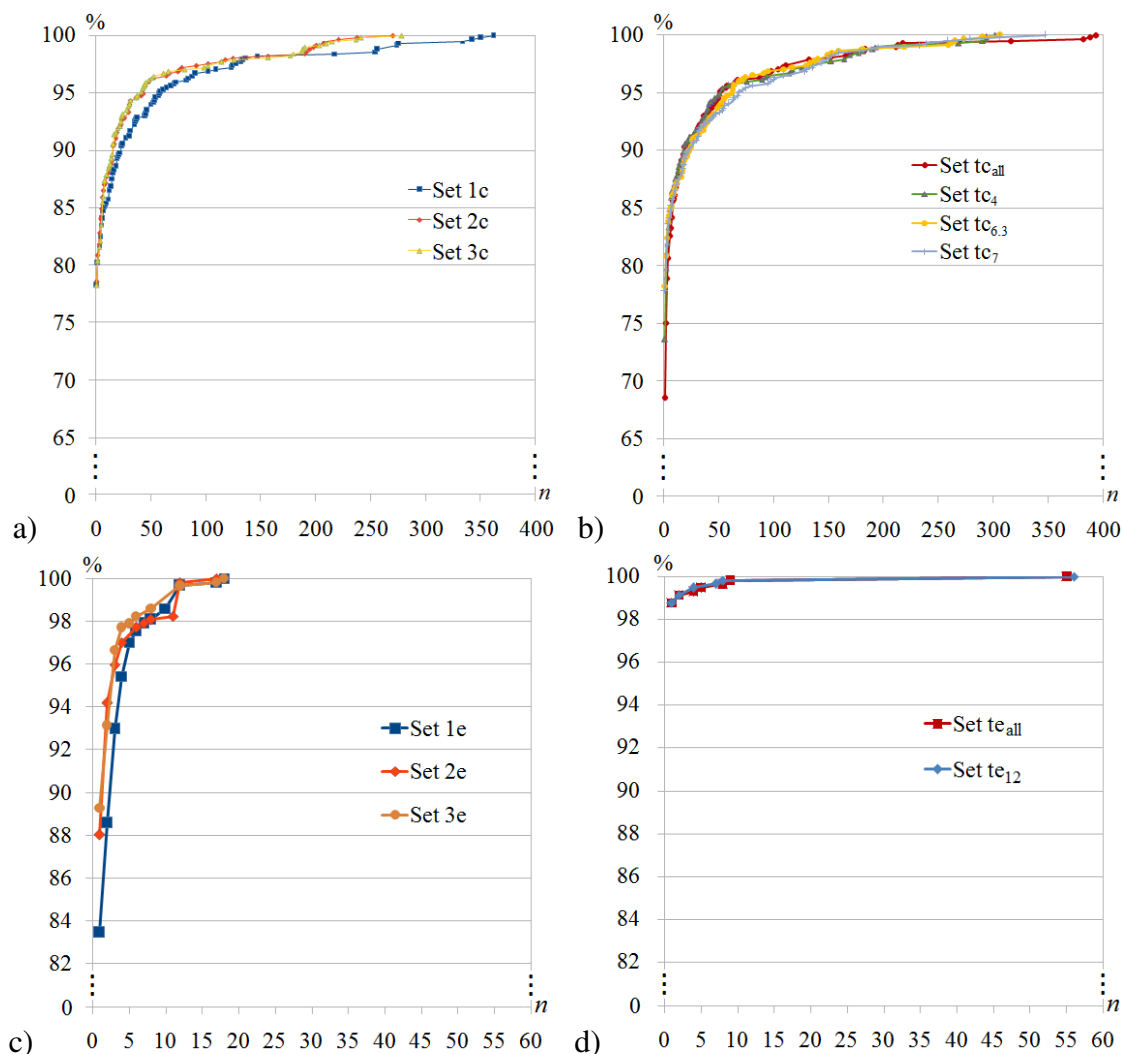
W następnym etapie dodano procedurę uczenia maszynowego. Procedurę zaimplementowano i przetestowano przy założeniu, że wiedza na temat poprzednich rozmów określonej osoby lub dotyczących określonego zdarzenia nie jest dostępna do nauki. Oznacza to, że informacje o dzwoniącym lub zdarzeniu ze zbioru testowego nie występowały w zbiorze treningowym. Ta właściwość jest ważna w typowych bazach CPR, gdzie ilość danych gwałtownie rośnie wraz ze wzrostem liczby nowych rozmówców i zdarzeń. Wyniki rozpoznawania rozmówcy przy wykorzystaniu zestawu parametrów wytrenowanych za pomocą procedury uczenia były podobne jak w przypadku manualnego doboru parametrów wyszukiwania, z tym, że dokładność rozpoznawania zdarzeń wzrosła do 99% (rys. 5b i 5d).

Dla zwiększenia precyzji rozpoznawania w niektórych szczególnych przypadkach, np. przy wykrywaniu oszusta na podstawie możliwości przemieszczenia się pomiędzy lokalizacjami, autor przetestował mechanizmy uwzględniające zależności pomiędzy cechami. Przy użyciu tych mechanizmów rozważane przypadki były modelowane dokładniej, a wyniki lepiej odzwierciedlały rzeczywiste niuanse.

Dodatkowo, rozpatrywany problem rozpoznawania został zdefiniowany jako problem klasyfikacji i rozwiązany przy użyciu sztucznej sieci neuronowej, za pomocą której dokonano klasyfikacji na dwa podzbiory, a mianowicie klasę rekordów należących do danego rozmówcy, grupy rozmówców, określonego zdarzenia lub grupy zdarzeń i klasę pozostałych rekordów. Rekordy reprezentujące klasy testowe uczestniczyły w etapie treningu. W przypadku rzeczywistego CPR odnosi się to do poszukiwania danej dzwoniącej osoby, która występuje na wcześniej zdefiniowanej tzw. czarnej liście osób, np. podejrzewanych o nadużywanie telefonu alarmowego. Skuteczność klasyfikacji przez sieć neuronową okazała się lepsza dla zrównoważonego zestawu reprezentantów obu klas. Jeśli liczba reprezentantów była niewystarczająca to dokładność klasyfikacji była niższa. Zatem, sieć neuronową należy stosować jedynie na ograniczonej grupie rekordów.

Podsumowując, można stwierdzić, że zaproponowany przez autora mechanizm wyszukiwania, dedykowany do wspierania operatorów telefonicznego numeru alarmowego w identyfikacji osób i zdarzeń na podstawie punktacji ważonych metadanych, najlepiej nadaje się do wstępnego sortowania wszystkich rekordów rozmów. Tym samym znacznie ogranicza listę rekordów do analizy. Dzięki temu operator analizuje tylko część z ogromnej bazy danych lub może dodatkowo skorzystać z sieci neuronowych do klasyfikacji wybranej grupy rekordów i znalezienia końcowego wyniku. Obszar poszukiwań operatora jest zawężony,

a zagrożenie jest odpowiednio oznaczone i zidentyfikowane. Proponowane metody zwiększą wydajność pracy i funkcjonalność stanowiska pracy nowoczesnych operatorów CCTV i CPR.



Rys. 5. Procent poprawnie rozpoznanych mówców (a i b) i zdarzeń (c i d) w funkcji miejsca na wynikowej liście rankingowej dla zestawów wag dobranych manualnie (a i c) i wytrenowanych przy pomocy procedury uczącej (b i d)

4. Wnioski

Niniejsza rozprawa doktorska została poświęcona interfejsom człowiek-komputer wspomagającym automatyczne rozpoznawanie zagrożeń. Autor zaproponował wizualizację stereowizyjną dla operatorów monitoringu wizyjnego oraz mechanizm rozpoznawania zdarzeń i osób dla operatorów systemu telefonicznego numeru alarmowego.

W zakresie systemów wizyjnych autor ustalił, że wystarczającą skuteczność uzyskują już metody wywoływania wyważonego efektu 3D bazujące na uproszczonych, dwuwartościowych mapach głębi. Ponadto wykazał, że pomiędzy parametrami sterującymi efektem 3D występują liniowe zależności, dzięki czemu w najprostszym przypadku liczbę parametrów można zredukować do jednego. Wyniki badań wskazują, że wprowadzenie efektu

stereowizyjnego pozwala na lepsze niż w przypadku obrazów 2D rozpoznawanie zagrożeń przez operatora monitoringu wizyjnego.

W zakresie telefonicznego numeru alarmowego autor zaproponował zbiór cech opisujących zgłoszenia oraz metodę ich porównań. Zaimplementowany system wspomaga rozpoznawanie powtarzających się zgłoszeń tych samych osób lub wielu zgłoszeń dotyczących jednego zdarzenia. Mimo występowania w rekordach jedynie niepełnych, niepewnych lub niepoprawnych danych uzyskano wysoką skuteczność rozpoznawania. Opracowano narzędzie, które znacząco zawężając obszar poszukiwań, ułatwia pracę operatora i przyspiesza podejmowanie decyzji.

Cel naukowy, którym był rozwój zautomatyzowanych mechanizmów przetwarzania informacji wizyjnej i audio do wspomagania operatorów centrów informacyjnych wraz z rozwojem odpowiedniego oprogramowania, został zrealizowany, a teza naukowa została udowodniona.

Bibliografia

- [Alla2007] W. Allasia, F. Gallo, F. Chiariglione, F. Falchi, *An innovative approach for indexing and searching digital rights*, Proc. of 3rd International Conference Automated Production of Cross Media Content for Multi-Channel Distribution, Barcelona, Spain, 28–30 November 2007, pp. 147–154
- [Balc2008] J. Balcerek, A. Dąbrowski, A. Konieczka, *Simple efficient techniques for creating effective 3D impressions from 2D original images*, Proc. of New Trends in Audio and Video/ Signal Processing: Algorithms, Architectures, Arrangements, and Applications (NTAV/SPA), IEEE International Conference, Poznań, Poland, 25–27 September 2008, pp. 219–224
- [Balc2009] J. Balcerek, S. Drgas, A. Dąbrowski, A. Konieczka, *Prototype multimedia database system for registration of emergency situations*, Proc. of Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), IEEE International Conference, Poznań, Poland, 24–26 September 2009, pp. 144–148
- [Balc2010] J. Balcerek, P. Pawłowski, A. Konieczka, S. Drgas, A. Dąbrowski, M. Kmieciak, *Database of emergency telephone calls – system tools for real-time registration and metadata searching*, Proc. of Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), IEEE International Conference, Poznań, Poland, 23–25 September 2010, pp. 89–94
- [Balc2011a] J. Balcerek, P. Pawłowski, *Multicriteria metadata mechanisms for fast and reliable searching of people using databases with unreliable records*, Proc. of 4th International Conference Multimedia Communications, Services and Security, MCSS, Krakow, Poland, 2–3 June 2011, Springer. Communications in Computer and Information Science, pp. 225–232

- [Balc2011b] J. Balcerek, A. Konieczka, A. Dąbrowski, T. Marciniak, *Binary depth map generation and color component hole filling for 3D effects in monitoring systems*, Proc. of Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), IEEE International Conference, Poznań, Poland, 29–30 September 2011, pp. 138–143
- [Balc2012] J. Balcerek, A. Konieczka, A. Dąbrowski, M. Stankiewicz, A. Krzykowska, *Approach to evoking stereovision impressions from images*, Electrical Review (Przegląd Elektrotechniczny), no. 6/2012, pp. 17–23
- [Balc2013] J. Balcerek, A. Dąbrowski, A. Konieczka A., *Stereovision option for monitoring systems – a method based on perception control of depth*, Proc. of Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), IEEE International Conference, Poznań, Poland, 26–28 September 2013, pp. 226–230
- [Balc2014] J. Balcerek, A. Dąbrowski, A. Konieczka, *Supporting stereovision tool for visual monitoring operators*, Electronics – Constructions, Technologies, Applications (Elektronika – konstrukcje, technologie, zastosowania), no. 4/2014, pp. 9–13
- [Balc2014a] J. Balcerek, A. Dąbrowski, A. Konieczka, *Fast 2D to 3D image conversion based on reduced number of controlling parameters*, Electrical Review (Przegląd Elektrotechniczny), no. 12/2014, pp. 11–16
- [Balc2015] J. Balcerek, P. Pawłowski, A. Dąbrowski, A. Konieczka, *Model of caller and event recognition mechanism for emergency telephone system* (in Polish), Proc. of 14th National Conference of Electronics (XIV Krajowa Konferencja Elektroniki), Darłówko Wschodnie, Poland, 8–12 June 2015, pp. 572–577
- [Balc2015a] J. Balcerek, P. Pawłowski, A. Dąbrowski, A. Konieczka, *Relations between features for automatic recognition of abnormal cases in emergency telephone call systems*, Proc. of Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), IEEE International Conference, Poznań, Poland, 23–25 September 2015, pp. 191–196
- [Bart2008] D. Bartz, D. Cunningham, J. Fisher, C. Wallraven, *The role of perception for computer graphics*, Eurographics 2008 – State of the Art Reports, The Eurographics Association, 2008
- [Berr2003] S. Berrani, L. Amsaleg, P. Gros, *Robust content-based image searches for copyright protection*, Proc. of 1st ACM International Workshop on Multimedia Databases, New Orleans, LA, USA, 7 November 2003, pp. 70–77
- [Bota2009] S. Bota, S. Nedevschi, M. Konig, *A framework for object detection, tracking and classification in urban traffic scenarios using stereovision*, Proc. of IEEE 5th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 27–29 August 2009, pp. 153–156
- [Buch2013] B. Buchwald, *Emergency Notification Center – Integrated Dispatcher System* (in Polish), Neokart Gis, Warsaw, Poland, available: www.neokartgis.com.pl, accessed on: 4 November 2013

- [Cetn2012] D. Cetnarowicz, A. Dąbrowski, M. Pleva, J. Juhar, S. Ondas, *Creation of event model in order to detect dangerous events*, D7.2, European Seventh Framework Programme, FP7-218086-Collaborative Project, The INDECT Consortium, 2012
- [Chan2012] Y.L. Chang, Y.P. Tsai, T.H. Chang, Y.R. Chen, S. Lei, *A depth map refinement algorithm for 2D-to-3D conversion*, Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012, pp. 1437–1440
- [Chen2012] Z. Chen, T. Ellis, S.A. Velastin, *Vehicle detection, tracking and classification in urban traffic*, Proc. of 15th International IEEE Conference on Intelligent Transportation Systems (ITSC), Anchorage, AK, USA, 16–19 September 2012, pp. 951–956
- [Dabr2010] A. Dąbrowski et al., *Biometric features analysis component based on video and image information*, D7.3, European Seventh Framework Programme, FP7-218086-Collaborative Project, 2010
- [Dabr2012] A. Dąbrowski, S. Drgas, P. Pawłowski, J. Balcerek, *Development of PUEPS corpus of emergency telephone conversations*, Proc. of Language Resources for Public Security Applications Workshop (LRPS) at 8th International Conference on Language Resources and Evaluation (LREC), Istanbul, Turkey, 27 May 2012, pp. 8–12
- [Dabr2012b] A. Dąbrowski, J. Balcerek, A. Konieczka, *An approach to adjustment and reduction of the number of controlling parameters for simple 2D to 3D image conversion schemes*, Proc. of New Trends in Audio and Video/ Signal Processing: Algorithms, Architectures, Arrangements and Applications (NTAV/SPA), IEEE International Conference, Łódź, Poland, 27–29 September 2012, pp. 227–231
- [Data2012] Data Sheet, *Bosch: IVA 5.50 Intelligent Video Analysis*, available: http://stna.resource.bosch.com/documents/Data_sheet_enUS_1559037707.pdf, accessed on: December 3, 2012
- [Drga2015] S. Drgas, A. Dąbrowski, *Speaker recognition based on multilevel speech signal analysis on Polish corpus*, Multimedia Tools and Applications, vol. 74, issue 12, 2015, pp. 4195–4211
- [ECCC2015] European Commission, Communications Committee, *Implementation of the European emergency number 112 – Results of the eight data-gathering round*, Working Document, Belgium, Brussels, 2015
- [EurC2011] Official website of the European Commission (EC), *Cities of the future – challenges, visions, perspectives*, available: http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/citiesoftomorrow/citiesoftomorrow_final_pl.pdf, 2011, accessed on: 4 March 2016
- [EurC2014] Official website of the European Commission (EC), *Crime statistics*, available: http://ec.europa.eu/eurostat/statistics-explained/index.php/Crime_statistics, 2014, accessed on: 16 January 2016

- [Gant1999] V. Ganti, J. Gehrke, R. Ramakrishnan, *Mining very large databases*, Computer, vol. 32, no. 8, 1999, pp. 38–45
- [Grzy2014] J. Grzybowska, P. Jaciów, M. Ziółko, *Emergency phone call features association analysis*, Proc. of National Conference on Applications of Mathematics in Biology and Medicine (Krajowa Konferencja Zastosowań Matematyki w Biologii i Medycynie), Łochów, Poland, 23–27 September 2014, pp. 46–51
- [Huyn2011] Q. Huynh-Thu, M. Barkowsky, P. Le Callet, *The importance of visual attention in improving the 3D-TV viewing experience: overview and new perspectives*, IEEE Transactions on Broadcasting, vol. 57, issue 2, 2011, pp. 421–431
- [ITPC2012] IT Projects Center, *3/4 of notifications on emergency number 112 is false or unjustified* (in Polish), 1 June 2012, available: http://www.cpi.gov.pl/art,34_zgloszen_na_numer_alarmowy_112_jest_falszywa_lub_niezasadna,251,92.html, accessed on: 15 May 2015
- [KimD2007] D. Kim, M. Dongbo, K. Sohn, *Stereoscopic video generation method using motion analysis*, Proc. of 3DTV Conference, Kos Island, Greece, 2007, 7–9 May 2007, pp. 1–4
- [Klem2009] P. Klement, V. Snasel, *Anomaly detection in emergency call data – the first step to the intelligent emergency call system management*, Proc. of International Conference on Intelligent Networking and Collaborative Systems, Barcelona, Spain, 4–6 November 2009, pp. 226–232
- [Koni2014] A. Konieczka, J. Balcerek, A. Dąbrowski, *Iterative method of impulse noise removing from digital images* (in Polish), Electronics – Constructions, Technologies, Applications (Elektronika – konstrukcje, technologie, zastosowania), no. 4/2014, pp. 21–23
- [Koni2016] A. Konieczka, J. Balcerek, P. Pawłowski, *An effective method of HDR images generation for CCTV systems*, Electrical Review (Przegląd Elektrotechniczny), no. 9/2016, pp. 102–105
- [Krie2006] H.P. Kriegel, P. Kroger, P. Kunath, A. Pryakhin, *Effective similarity search in multimedia databases using multiple representations*, Proc. of 12th International Multi-Media Modelling Conference, Beijing, PRC, 4–6 January 2006, pp. 389–392
- [LinC2005] W.S. Lin, M.K. Chuang, G. Tien, *Autonomous mobile robot navigation using stereovision*, Proc. of IEEE International Conference on Mechatronics (ICM), Taipei, Taiwan, 10–12 July 2005, pp. 410–415
- [LiWa2010] S. Li, F. Wang, W. Liu, *The overview of 2D to 3D conversion system*, Proc. of IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design (CAIDCD), Yiwu, PRC, 17–19 November 2010, vol. 2, pp. 1388–1392
- [LuoK2007] X. Luo, R.V. Kenyon, D. Kamper, D. Sandin, T.A. DeFanti, *The effects of scene complexity, stereovision, and motion parallax on size constancy in a virtual environment*, Proc. of IEEE Virtual Reality Conference, Charlotte, NC, USA, 10–14 March 2007, pp. 59–66

- [McCa2002] M. McCahill, C. Norris C, *Urbaneye: CCTV in London*, Centre for Criminology and Criminal Justice, University of Hull, UK, June 2002
- [OBri2013] T. O'Brien, *112 and the EU Legislative Framework*, European Emergency Number Association (EENA) Operations Document, 2013
- [Pozn2016] Poznań City Hall, *Video monitoring system of Poznań* (in Polish), available: <http://www.poznan.pl/mim/main/system-monitoringu-wizyjnego-miasta-poznania,p,1443,1743,1744.html>, accessed on: 7 January 2016
- [Rank2012] S. Rankin, N. Cohen, K. Maclennan-Brown, K. Sage, *CCTV operator performance benchmarking*, Proc. of IEEE International Carnahan Conference on Security Technology (ICCST), Boston, MA, USA, 15–18 October 2012, pp. 325–330
- [Rash2009] U. Rashid, I. A. Niaz, M. A. Bhatti, *M3L: architecture for multimedia information retrieval*, Proc. of 6th International Conference on Information Technology: New Generations, Las Vegas, NV, USA, 27–29 April 2009, pp. 1067–1072
- [Rawa2010] A. Rawal, M.K. Kowar, S. Sharma, H.R. Sharma, *Automated document ranking evaluation in digital libraries*, Proc. of International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom), Kottayam, India, 16–17 October 2010, pp. 258–260
- [Reve2008] A. Reveiu, M. Dardala, I. Smeureanu, *A MPEG-21 based architecture for data visualization in multimedia web applications*, Proc. of International Conference Visualization (VIS), London, UK, 9–11 July 2008, pp. 84–89
- [Sici2014] M.A. Sicilia, *Handbook of metadata, semantics and ontologies*, World Scientific Publishing Co., Singapore, 2014
- [Stat2015] Statista, the statistics portal, *The ten most traffic jam prone cities in Europe*, 2015, available: <http://www.statista.com/statistics/235780/most-traffic-jam-prone-cities-in-europe/>, accessed on: 4 March 2016
- [Supr2013] Supreme Audit Office, *Preparation of the system of civil protection against natural disasters and emergencies – information on audit results* (in Polish), 20 June 2013, available: <http://www.nik.gov.pl/plik/id,5308,vp,6885.pdf>, accessed on: 15 May 2015
- [TamZ2006] W.J. Tam, L. Zhang, *3D-TV content generation: 2D-to-3D conversion*, Proc. of IEEE International Conference on Multimedia and Expo (ICME), 9–12 July 2006, Toronto, Canada, pp. 1869–1872
- [Urey2011] H. Urey, K.V. Chellappan, E. Erden, P. Surman, *State of the art in stereoscopic and autostereoscopic displays*, Proceedings of the IEEE, vol. 99, issue 4, 2011, pp. 540–555
- [Velt2012] J. Velten et al., *Preliminary multimodal features detection system prototype*, European Seventh Framework Programme, D7.6, FP7-218086-Collaborative Project, The INDECT Consortium, 2012

- [Wang2013] B. Wang, V. Fremont, *Fast road detection from color images*, Proc. of IEEE Intelligent Vehicles Symposium (IV), Gold Coast, Australia, 23–26 June 2013, pp. 1209–1214
- [Zhao2012] J. Zhao, S. Ma, W. Han, Y. Yang, X. Wang, *Research and implementation of license plate recognition technology*, Proc. of 24th Chinese Control and Decision Conference (CCDC), Taiyuan, PRC, 23–25 May 2012, pp. 3768–3773
- [Ziaj2011] K. Ziaja, *The project of algorithm of rescue notification in UML object notation* (in Polish), Zeszyty Naukowe WSZOP, Katowice, Poland, vol. 7, no. 1, 2011, pp. 101–122

W wykazie literatury wymieniono jedynie te pozycje, które zostały zacytowane w niniejszym streszczeniu.