prof. Andrzej Czyżewski

Politechnika Gdańska, Wydział ETI

e-mail: ac@pg.gda.pl

05-04-2017 r.

# Recenzja rozprawy doktorskiej

## Juliana Balcerka

## zatytułowanej

## „Human-computer supporting interfaces for automatic recognition of threats"

## („Interfejsy człowiek-komputer do wspomagania automatycznego rozpoznawania zagrożeń")

## 1. Problem badawczy i jego znaczenie:

Celem pracy było opracowanie automatycznych mechanizmów, które wspierałyby pracowników systemów monitoringu i telefonicznych centrów alarmowych. W tym kontekście wytworzone i przetestowane zostały algorytmy automatycznej konwersji obrazu 2D do obrazu anaglifowego 3D w celu ułatwienia oceny przestrzennych relacji między obiektami i przyspieszenia zliczania osób. Autor przytacza także inny cel tej wizualizacji: „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". Drugi rodzaj opracowanych metod i przeprowadzonych badań dotyczy usprawnienia przeszukiwania baz danych nagrań w systemach telefonicznego numeru alarmowego i w tym celu autor wprowadza miary podobieństw osób dzwoniących i raportowanych przez nie zdarzeń, metody porównania, przetwarzania danych niepewnych, tworzy obszerne bazy przykładowych nagrań i przeprowadza i dokumentuje eksperymenty.

Tematyka rozprawy jest aktualna i ważna, zaś jej Autor podejmuje zagadnienia nie w pełni rozwiązane, zatem niewątpliwie warte bliższego poznawania i rozwiązywania na gruncie naukowym i praktycznym.

**Teza naukowa:** „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 stara się udowodnić tę tezę przez zaproponowanie sposobów konwersji obrazów 2D na 3D i eksperymenty przy użyciu zarejestrowanych obrazów z monitoringu wizyjnego oraz wykorzystanie metod porównywania metadanych opisujących zdarzenia i mówców. W pracy brak ścisłych informacji o wdrożeniu wyników badań do eksploatacji w centrach informacyjnych, zatem wydaje się, że stwierdzenie zawarte w tezie brzmi powinno być sformułowane w sposób mniej kategoryczny.

W rozdziale 1 został opisany obszar badań, cele, teza naukowa i zakres badań. Problemy, które zostały rozwiązane w pracy, odniesiono do zastanego stanu wiedzy. Praktyczna część podzielona została na dwa główne obszary: rozdziały 2 i 3 dotyczące tworzenia i oceny obrazu 3D i rozdziały 4 i 5 opisujące metody przetwarzania metadanych opisujących rozmowy telefoniczne w systemach telefonu alarmowego. Pierwsza zawiera 59 stron, druga 66 stron. Nie jest jasna decyzja Autora o połączeniu w jednej rozprawie dwóch odmiennych dziedzin, tworzenia i badania metod przeznaczonych do różnych zastosowań, których częścią wspólną wydaje się być jedynie aspekt poprawy bezpieczeństwa. Każda z części, choć przedstawiona w postaci, która wymagała sporego nakładu pracy, nie jest wystarczająco rozbudowana, by stanowić treść samodzielnego tekstu rozprawy, a ich suma tworzy rozłączną konstrukcję dwuczęściową, bowiem analiza obrazu trójwymiarowego jest czymś całkowicie odmiennym od zagadnienia rozpoznawania mówców.

**Czy ma on charakter naukowy?**

Podjęte problemy są w istocie naukowe, jednakże metodyka oceny uzyskanych wyników miejscami nie w każdym przypadku jest kompletna. W rozdziale 2 opisane są mechanizmy percepcji, sposoby tworzenia i wizualizowania obrazu stereoskopowego. Autor rozwiązuje istotne praktyczne problemy - proponuje sposoby uzupełniania brakujących pikseli dla metody konwersji obrazu 2D do 3D. Następnie dokumentuje przeprowadzone testy subiektywnej oceny jakości efektu, zauważalności artefaktów, postrzegania głębi, jednakże sposób analizy danych nasuwa wątpliwości, które są opisane w trzeciej części tej recenzji. Autor wykazuje zależność liniową między wybranymi parametrami algorytmu, którą wykorzystuje do zredukowania liczby parametrów. Odzwierciedla to, zdaniem recenzenta, interesujący, pomysłowy i przydatny kierunek poszukiwań. W rozdziale 3 autor dokumentuje przeprowadzone subiektywne testy oceny głębi, zależności przestrzennej między obiektami, trafności w liczeniu osób w obrazie ale ponownie zaproponowaną metodykę trzeba ocenić, jako dyskusyjną w pewnych szczegółach.

W rozdziale 4 i 5 Autor definiuje i podaje przykłady metryk odległości dla trzech typów zmiennych: binarnej, ciągłej, kategoryzacyjnej i wprowadza miarę korelacyjną. Autor proponuje ich zastosowanie do porównywania wielowymiarowych rekordów z bazy danych

zawierającej metadane rozmów rejestrowanych w systemie telefonu alarmowego. Postawiony problem ma charakter naukowy, jednak ponownie zaproponowana metodyka i sposób prowadzenia testów i analizy ich wyników pozostawiają pewien niedosyt.

**Czy ma on znaczenie praktyczne?**

Problemy tworzenia i prezentacji materiału stereoskopowego są istotne i mają liczne liczne praktyczne zastosowania, nie tylko w monitoringu, ale w szeroko rozumianych multimediach: w rozrywce, wizualizacji architektonicznej, medycznej i naukowej. Również metody porównywania i przeszukiwania wielowymiarowych metadanych w celu oceny podobieństw między cechami mówców i atrybutami rejestrowanych zdarzeń są wartościowe z praktycznego punktu widzenia. Potencjalny wymiar praktyczny korzystnie wpływa na jej ogólna wartość.

## 2. Wkład autora

Autor w rozprawie powołuje się na 18 współautorskich publikacji z lat 2008 – 2016, w tym 3 w Przeglądzie Elektronicznym, 2 w czasopiśmie Elektronika, 1 rozdział w książce i 12 referatów konferencyjnych. 10 publikacji dotyczy dziedziny stereowizji, 8 analizy metadanych w bazach nagrań. Z braku dokładnych danych recenzent zakłada, że własny wkład autora w tych publikacjach związany jest z kluczowymi metodami: tworzenia obrazu 3D i metody porównań metadanych. W takim ujęciu dorobek publikacyjny Autora należy uznać za istotny i dobrze powiązany tematycznie z przedmiotem rozprawy.

Problem tworzenia obrazów anaglifowych obecny jest w fotografii i kinie od przełomu XIX i XX wieku. Metody konwersji 2D do 3D z wykorzystaniem mapy głębi zostały już zaimplementowane i są wykorzystywane zarówno w aplikacjach działających w czasie rzeczywistym (np. odtwarzacze wideo dokonują na bieżąco automatycznej konwersji i prezentują filmy w pseudo-3D), jak edytorach wideo z możliwościami ręcznej edycji szczegółów. Istotny wkład autora dostrzegalny jest w sposobie ograniczania liczby parametrów algorytmu – Autor wykazał występowanie liniowych zależności między parametrami i możliwość zastosowania tylko jednego do generowania obrazów anaglifowych 3D o dobrej jakości.

Wartościowy własny wkład Autora w dziedzinę analizy wielokryterialnej, metod i metryk porównywania danych dotyczy praktycznego zastosowania odpowiednich narzędzi: samodzielnego przygotowania obszernej bazy nagrań wideo inscenizowanych zdarzeń, nagrań fonicznych rozmówców symulowanego telefonu alarmowego oraz zgromadzenia metadanych setek nagrań i propozycji metodologii porównania, zastosowania miary globalnego podobieństwa, propozycji dostrajania wag ręcznie i z zastosowaniem sieci neuronowej.

## 3. Poprawność

Wymagana ocena poprawności rozprawy jest w pewnym stopniu utrudniona, ze względu na dość liczne uwagi polemiczne, odnoszące się do treści rozprawy. Przedstawione w rozprawie stwierdzenia, metody i sposoby analizy wyników w pewnych przypadkach nie są bezbłędne.

Najważniejszy niedosyt został spowodowany niestosowaniem przez Autora narzędzi statystycznych oprócz bardzo prostych, tzn. histogramu i wartości średniej. Stawiane tezy (główna i pomocnicze) są udowadniane, ale istotność statystyczna uzyskiwanych wyników nie jest jednocześnie analizowana. W szczególności, recenzent pragnie poznać odpowiedzi Autora na zestawione poniżej kwestie polemiczne:

1. W zagadnieniu tworzenia obrazów 3D Autor nie wspomina o istniejących wariantach obrazów anaglifowych: anaglif kolorowy, pół-kolorowy, Dubois. Pomija istotne problemy rywalizacji siatkówkowej i zaburzeń kolorów, które utrudniają odbiór obrazu anaglifowego, powodują szybkie zmęczenie, nie pozwalają ocenić prawdziwych barw przedmiotów, co w monitoringu jest kluczowe. Celem aplikacji było dłuższe niż przy klasycznym obrazie 2D utrzymywanie uwagi operatora monitoringu, jednak nierozwiązane istotne wady techniki anaglifowej pozostawiają pewne wątpliwości odnośnie celowości użycia w podanym kontekście.
2. Autor nie pisze skąd w realnych zastosowaniach dla jednego strumienia 2D z kamery monitoringu można pozyskać mapę głębi, która ma posłużyć do jego konwersji do postaci 3D.
3. Autor, bez podania przekonującego uzasadnienia, stosuje mapę głębi binarną, zamiast wielostopniowej. Metoda użyta przez Autora nie umożliwia prostego rozszerzenia na mapy wielostopniowe.
4. Przykład działania metody "segment shift" przedstawiony jest na całym obrazie, tzn. obiekt pierwszoplanowy i tło nie są rozdzielane i oba przesuwane są o ten sam dystans (rys. 2.14). Tymczasem przykład prostej konwersji (rys. 2.11) wykorzystuje mapę głębi i stosuje inne przesunięcie do tła i inne do obiektu. Porównanie wyników obu metod jest wobec tego niemożliwe. Analogiczna uwaga dotyczy przykładu działania metody "segment scaling" (rys. 2.15).
5. W kwestionariuszu testowym (str. 37) występuje pytanie "How does the distance between the viewer and the displayed image change", co sugeruje, że percypowana zmiana dystansu ma być oceniona w stosunku do referencyjnego obrazu. Tymczasem Autor nie wspomina o możliwości wyświetlania naprzemiennie obrazu z głębią i bez głębi, więc ocena zmiany odległości jest niemiarodajna w tym wypadku. Dopiero analiza zawartości rys. 2.19 ujawnia, że w GUI aplikacji testowej pytanie brzmi "Czy obraz zagłębia się w monitor czy wychodzi z niego? Bliżej / Głębiej".
6. W przypadku pytania: "What is the perceptible quality of the picture?" zamiast „good", „medium", „bad" lepiej byłoby skorzystać z miar degradacji jakości z rekomendacji ITU-R BT.500-11 "Methodology for the subjective assessment of the quality of television picture": 5 = imperceptible, 4 = perceptible, but not annoying, 3=slightly annoying, 2=annoying, 1=very annoying.
7. Na wykresie 2.20 i kolejnych pionowa oś wyskalowana jest w procentach liczby osób, które widzą efekt 3D i zbliżanie lub oddalanie obiektu. Każda próbka jest wynikiem średniej ocen pozyskanych od 71 osób i przy tak licznej grupie powinna być wykonana analiza statystyczna tych wyników, mediany, prezentacja kwartyli, wyznaczenie istotności np. dla hipotezy o równości średnich. Średnia jakość obrazu, rys. 2.22. także powinna być przeanalizowana dokładniej i wykreślona np. w formie wykresów pudło-wąsy. Teza, że

postrzegana jakość obrazu silnie zależy od treści obrazu (str. 40) powinna być udowodniona w toku analizy statystycznej.

8. Złudzenie przybliżania obrazu do widza wydaje się nie zależeć od kierunku przesunięcia (rys. 2.23) co jest sprzeczne z założeniami o dodatniej i ujemnej paralaksie w każdej metodzie prezentacji obrazu 3D. Autor nie komentuje tego faktu i wyników pomiaru.
9. W pytaniu o odległość uczestnicy mają do wyboru „Bliżej" lub „Głębiej" (sic). Korzystne dla przejrzystości wizualizacji tych obserwacji byłoby prezentowanie na jednym wykresie wartości dla bliżej, dalej i „nie widzę efektu 3D", sumujące się do 100%. Tymczasem, około 30% osób odpowiada, że dana konfiguracja (zwłaszcza przesunięć w prawo) skutkuje oddaleniem i tyle samo osób, że przybliżeniem do widza. Recenzent chciałby poznać interpretację takiego wyniku.
10. Na rys. 2.27 warto byłoby wykreślić linię trendu i wartość współczynnika korelacji, co wsparłoby tezę (str. 45), że jeśli widz zauważa mniej artefaktów to wystawiana przez niego ocena jakości jest wyższa.
11. Rozdział 2.8.1. Autor kończy pisząc, że metoda daje stabilny, silny, akceptowalny efekt 3D, czego niestety wcześniejsze wyniki jednak nie potwierdzają. Najmniejsze problemy (mylenie przesunięcia obraz w przód i w tył, częste odpowiedzi „nie widzę efektu 3D", artefakty widziane przez 30%-70%) występują, bowiem, dla wartości przesunięcia -4,5%.
12. Wykresy 2.31 – 2.34 zawierają wiele próbek ulokowanych na lub poniżej prostej y=x. To sugeruje, że uczestnicy testu ustalali głębię obiektu i tła identycznie, bez brania pod uwagę różnicy głębi między nimi. Autor powinien zamieścić komentarz o tym, że brak różnic głębi był nieświadomie preferowany przez bardzo wiele osób.
13. Eksperyment w podrozdziale "Dependence of linear parameters", str. 54, jest niedostatecznie udokumentowany. Zadaniem uczestnika było dostosowanie wartości jednego parametru, którego preferowanych wartości Autor nie pokazuje. Interesujące byłoby odniesienie się do poprzednich eksperymentów z rozdziału 2.8.1. i sprawdzenie, czy potwierdza się optymalna wartość przesunięcia o -4,5%.
14. W eksperymencie 2.8.3. Autor niepoprawnie zakłada, że bez względu na odległość obiektu od widza można stosować tą samą liniową funkcję do wyznaczenia przesunięcia kanału czerwonego. Powoduje to zaburzenie efektu 3D, gdyż wprowadza stałą różnicę głębi między tłem a obiektem. Tymczasem dla ujęć nazwanych "Image2, 3, 4" obiekt jest bardzo daleko od widza i względna różnica między jego odległością a tłem powinna być o wiele mniejsza niż dla ujęć bliskich: "image1, 5".
15. Kluczowe eksperymenty w rozdziale 3 wykonywane są na ujęciach, które nie są typowe dla kamer monitoringu, z kamerą na statywie na poziomie < 2m, oś zorientowana horyzontalnie. Tymczasem w monitoringu dominują kamery na słupach, pod sufitem, czyli >=2,3m i osie skierowane w dół. Przykłady niewłaściwych ujęć na rys. 2.10, 2.18, 2.29, 3.2, 3.3, 3.4, 3.5, 3.6, 3.8, 3.9, 2.10. Przykłady właściwych ujęć rys. 2.36, 2.39, 2.40, 3.1.
16. Test w rozdziale 3.1.1. "Simplified depth maps" (str. 62) polega na wykryciu, że osoba stojąca dalej niż obiekty pierwszoplanowe w sposób sztuczny została przesunięta przed płaszczyznę ekranu, bliżej niż obiekty pierwszoplanowe. Przydałoby się wyjaśnienie, jakie uzasadnienie praktyczne ma operacja przybliżania obiektu dalekiego przed ekran, silnie zaburzająca naturalne percypowanie głębi obrazu?

17. Przykłady na rys. 3.3a,b, 3.4a,b zawierają wiele pikseli oryginalnie w kolorze czerwonym, co, w trakcie obserwacji przez okulary anaglifowe powoduje silną rywalizację siatkówkową (ang. retinal rivalry). Tymczasem Autor nigdzie w swojej pracy nie odnosi się do tego znanego problemu, utrudniającego percepcję obrazów anaglifowych, powodującego szybkie zmęczenie i duży dyskomfort widza.
18. Percypowany dystans między dwiema osobami w eksperymencie 3.1.3 (str. 65) wyrażony jest tylko jako wartość uśredniona, nie dając wyobrażenia o problemach uczestników testu z oceną tej odległości. Jaka jest wariancja w odpowiedziach uczestników? Ocena odległości dla obrazu 2D i 3D wymaga analizy statystycznej i wykazania istotności różnic. Podobna uwaga o wykorzystywaniu tylko średniej dotyczy wszystkich wyników w rozdziale 3.
19. Ocena bezpośredniego kontaktu między osobami/przedmiotami w rozdziale 3.1.4 polega wyłącznie na obserwacji bliskich ujęć. Autor nie komentuje faktu, że dla obiektów znajdujących się dalej spadnie przestrzenna rozdzielczość oceny odległości i niemożliwa będzie ocena kontaktu osób dalekich od kamery. Z praktycznego punktu widzenia wspieranie monitoringu w ten sposób nie jest przekonujące, gdyż nie zawsze w bezpośredniej bliskości kamer dochodziło będzie do istotnych zdarzeń. W rozdziale 3.1.4 użyto 2 próbki pozytywne i 4 negatywne i taki brak balansu zmniejsza istotność testu. Ponadto jedyna próbka 3D pozytywna jest oceniona przez 87% jako negatywna, co w jakimś stopniu podważa tę metodę w zastosowaniu do stwierdzania istnienia kontaktu między osobami, gdyż z testu wynika, że zbyt duży odsetek widzów zawsze wskazuje na brak kontaktu.
20. W rozdziale 3.1.5 w ocenie relacji trajektorii między obiektami dla wielu próbek trafność jest losowa tj. 50-procentowa, a dla innych przeważa ocena nieprawidłowa, nawet do 72%. Technika 3D nie nadaje się do tego typu zadań, czego Autor nie komentuje. Autor pisze, że metoda wymaga, aby obiekty widoczne były obecne w kadrze w całości, jednak żadna z ocenianych próbek nie przedstawia takiego przypadku, więc taka hipoteza nie jest uprawniona w kontekście materiału przedstawionego w pracy. Skuteczna ocena relacji trajektorii prawdopodobnie wymagałaby raczej obserwacji filmu, a nie zdjęć. Podobnie test liczenia osób na statycznych zdjęciach jest zadaniem niezwiązanym bezpośrednio z praktycznym monitoringiem wideo.
21. Tabela 3.7 i 3.8 w sposób nieczytelny prezentuje wyniki. Liczba osób 10 i „correct counting" równe 16.7% oznacza, że osoby średnio zliczyły 1.67 osoby czy 16.7% widzów zliczyło dokładnie 10 osób? Jakie były wariancje tych zliczeń? Średnia procentowa jest niezrozumiała w tym kontekście.

W dziedzinie **analizy metadanych** najbardziej zauważalne potknięcia także związane są z niedostateczną uwagą poświęconą analizie wyników testów i z niezastosowaniem lub niewłaściwym stosowaniem kluczowych metod drążenia danych.

22. Na stronach 77-78 Autor opisuje sposób dyskretyzacji wartości dystansu, jednak ani razu nie używa terminu „dyskretyzacja". Niektóre przedziały mają szerokość 5, inne 10 i 40. Zastępowane są wartościami dyskretnymi w nieregularnych interwałach: 0, 0,1, 0,3, 0,5, 0,6, 0,7, 0,8, 1,0. Zaproponowane arbitralnie przedziały wartości ciągłej i odpowiadające

im wartości dyskretne nie zostały przez Autora w żaden sposób uzasadnione. W literaturze i w praktyce częste są metody dyskretyzacji, takie jak: Maximum Discernability, lub maksymalizacja zysku informacyjnego (ang. Information Gain), do których Autor się nie odnosi, ani nie stosuje ich. Taka sama uwaga dotyczy braku uzasadnienia dla sposobu przypisania prawdopodobieństw do MTT (średni czas podróży) (str. 80) oraz braku uzasadnienia dla przedziałów wiekowych (str. 82). Ponownie Autor charakteryzuje te wartości w rozdziale 5.2. (str. 111) uzasadniając zdawkowo: „they have been calculated on the basis of the real data or determined on the basis of the author's experience for other comparison types".

23. Niewłaściwe jest stosowanie nomenklatury dotyczącej obrazu do wartości w macierzy (str. 85): „[...] number of the shadowed boxes rises. [...]boxes are just become to be shadowed or separated from the others [...] contrast of the whole image starts to decrease, down to the totally black image [...]". Autor powinien pisać o zmniejszających się różnicach wartości i o zwiększaniu się wartości w przedziale <0, 1>. Pojęcie szarości i kontrastu nie ma wprost przełożenia na wizualizowany w tabeli współczynnik korelacji. Podobnie na str. 93 Autor komentuje zwizualizowane macierze o coraz większych rozmiarach „increasing number of features, the image resolution is increasing". Rozdzielczość obrazu to liczba pikseli na cal (np. na ekranie lub na wydruku), co nie ma wprost przełożenia na rozmiar macierzy zawierającej dane arbitralne, nie będące wartościami koloru i nie prezentujące w tym wypadu pikseli obrazu. Uprawnione jest w tym wypadku pisanie wyłącznie o zwiększającym się rozmiarze macierzy.
24. Autor proponuje wyznaczanie korelacji wielu cech jednocześnie przy użyciu macierzy wielowymiarowych. Jednakże nie jest uzasadnione, ani jasne dlaczego w Tabeli 4.8 (str. 91) wartości v1 do v5 powtarzają się i wypełniają 25 komórek, bez podanych nazw kolumn i wierszy. Wektor pięcioelementowy, z kolejnymi wartościami v1 do v5 przekazywałby te same informacje. Dlaczego w Tabeli 4.9 (str. 92) występuje wiersz Śrem-Głogowska, skoro ulicy o takiej nazwie nie ma w tym mieście? Recenzent nie dostrzega konieczności używania wszystkich, nawet nieistniejących wartości lingwistycznych, gdyż prowadzi to do niepotrzebnego zwiększania wymiaru każdej takiej tabeli. Autor pisze, że w 3 miastach występuje 2095 unikatowych nazw ulic, ale np. 1937 w Poznaniu i 144 w Śremie. Zamiast dla podróży z Poznania do Śremu tworzyć tabelę 2092 na 2092 wystarczy więc tabela 1937 na 144. Podobna uwaga dla Tabeli 4.10.
25. Na etapie treningu (rozdział 4.4, str. 93) Autor zakłada, że cechy są niezależne, tymczasem wcześniejsze przykłady pokazują, że np. dana ulica w jednym mieście występuje a w innym nie, wobec czego zależność istnieje. W dalszej części tego rozdziału Autor prezentuje sposób określania wag dla globalnego wyniku, który definiowany jest 9 stron wcześniej. Warto byłoby przedstawić go ponowie w tym miejscu. Autor zakłada bez dodatkowego wyjaśnienia, że waga dobierana jest z przedziału <0,30>, co oznacza, że globalny wynik może przyjąć wartości od 0 do wartości o wiele przekraczających 100. Interpretacja wartości bezwzględnej wyniku „global score" nie jest podana. Autor jednym zdaniem pisze (str. 95), że metoda zastosowana może być także do trenowania wag korelacji między cechami, jednak nigdzie w opisie miar korelacyjnych Autor nie wprowadził wartości wag $w_{fl,fk}$. Wobec tego intencja jest niejasna i nie podano przykładu, który poparłby stwierdzenie, że metodę stosować można w taki sposób. Ponadto na etapie

eksperymentu (str. 114) Autor niejako wycofuje się z zaproponowanej w rozdziale 4.4. metody treningu wag i prezentuje wyniki dla wag ustalonych ręcznie i dopiero w rozdziale 5.3. przedstawia wyniki z użyciem treningu wag. Zbiór dzieli w nietypowych proporcjach, do treningu wykorzystując tylko 15% przypadków. Interesujące byłoby zweryfikowanie skuteczności w k-krotnej kros-walidacji. Recenzent pragnie przypomnieć, że 10-krotna kros walidacja stosuje trening na 100-10=90% zbioru. Z tego względu, że zwykle ponad 50% zbioru używana jest do treningu, ten opisywany wynik metody trenowanej tylko 15 procentami przypadków wydaje się nie do końca wiarygodny.

26. Nie przeprowadzono żadnej analizy danych, która wykazałaby, że przyjęte czasy podróży i miejsca docelowe wiążą się ze sobą z podawanymi przez Autora prawdopodobieństwami. Autor nie wyjaśnia, w jaki sposób możliwe byłoby pozyskanie wiarygodnych wartości tych prawdopodobieństw. Autor przedstawił metody, jednakże ich zastosowania pozostają hipotetyczne, tak długo, dopóki nie zostanie opracowana i zweryfikowana metodyka wyznaczania wspomnianych prawdopodobieństw.
27. W rozdziale 4.5. Autor proponuje zastosowanie sieci neuronowej o dwóch wyjściach do binarnej klasyfikacji przynależności rekordu z bazy danych rozmów do zadanej klasy. Klasy te opisane są następująco (str. 95: „The first class is the group of conversations which meet certain searching criteria. The second class contains those conversations, which do not meet these criteria."). Tzn. konkretna sieć jest nastrojona na jeden typ wyszukiwania z góry określony. Podejście to wydaje się niepraktyczne, gdyż nie porównuje rozmów między sobą, a poszukiwanie wg innych kryteriów wymaga użycia innej sieci. W rozdziale 5.5.1 Autor definiuje 6 klas, trenuje i testuje 6 różnych sieci. Klasy są dobrane arbitralnie: np. młode osoby dzwoniące przed południem, młode osoby mówiące szybko, interwencje z 20 marca. Zastosowanie dla innego kryterium wymaga definiowania i treningu nowej sieci, co jest niepraktyczne i znacząco utrudniłoby stosowanie rozwiązania przez pracownika telefonicznego centrum alarmowego, który miałby prowadzić taki trening algorytmu decyzyjnego. Uniwersalne podejście mogłoby polegać chociażby na równoległym podaniu na wejście sieci przypadku referencyjnego i przypadku nieznanego, a sieć generować powinna na jednym wyjściu miarę podobieństwa osób dzwoniących, a na drugim wyjściu miarę podobieństwa opisu zdarzenia. W rozdziale 5.5.2. proponuje sieć do podobnego zadania, jednak trenuje ją i testuje atrybutami nagrań jednej osoby, sieć ograniczona może być więc do działania na wyłącznie jednej osobie. Recenzent pozwala sobie zasugerować alternatywne podejście w którym sieć jest czarną skrzynką umiejącą porównywać dwa wektory podane równolegle na wejścia, a nie „czarną skrzynką" rozpoznającą wyłącznie konkretną osobę.
28. Rys. 5.20 prezentuje efektywności w zależności od wielkości zbioru uczącego. Do treningu użyto aż do 100% wszystkich przypadków, co jest niewłaściwe, bo prowadzić może do przetrenowania sieci, powoduje, że trening i test muszą być wykonane na tych samych danych i wyniki stają się mniej wiarygodne.
29. W opisie eksperymentu, str. 102, Autor pisze o potencjalnych trudnościach z niedokładnością danych i błędami. Wobec tego należy unikać przydzielania prawdopodobieństw o wartościach 0 różnym zdarzeniom nawet niemożliwym, gdyż mogą one wynikać z błędu ludzkiego. Destrukcyjne będzie dla dalszej analizy wykluczanie przypadku o zerowym prawdopodobieństwie.

30. Na rys. 5.5 rekordy mają identyczny typ i kategorię a dodatkowo rekord pierwszy-referencyjny i trzeci mają ten sam adres. Dlaczego wobec tego rekord drugi o innym adresie ma większe podobieństwo do referencyjnego? Autor nie komentuje zawartości tego rysunku, ani nie podaje przykładów wyliczania podobieństw między przykładowymi rekordami.
31. W tabeli 5.6 i 5.8. występują prawdopodobieństwa równe 1, 0,1 i 0. Tymczasem niekonsekwentnie w opisie poniżej Autor podaje, że przyjął binarne wartości. Tabele 5.3 do 5.8. zawierają arbitralnie przyjęte miary, których wartości niestety Autor nie uzasadnia inaczej, jak tylko swoim doświadczeniem. Kluczowe dla naukowej oceny poprawności metody jest udowodnienie, że odpowiednie wartości stałych, wag, progów bądź prawdopodobieństw, które występują w rozprawie, rzeczywiście generują duży np. zysk informacyjny (wyrażany jako spadek entropii następujący w wyniku przeprowadzenia dyskretyzacji daną wartością progową). Metod obiektywnej oceny skuteczności modelu Autor nie opisał.
32. W rozdziałach 5.4.1 i 5.4.2 Autor podaje możliwość rozpoznawania prawdziwych i nieprawdziwych komunikatów w rozmowach telefonicznych, potwierdza to zaledwie na pojedynczych przykładach. Tego typu tezy wymagają dokładniejszych eksperymentów na statystycznie istotnej populacji przypadków.

**Aspekty wartościowe rozprawy, to:** istotny wkład autora w analizę liniowych zależności między parametrami i możliwość zastosowania tylko jednego parametru do generowania obrazów anaglifowych 3D o dobrej jakości oraz samodzielne przygotowanie obszernej bazy nagrań wideo inscenizowanych zdarzeń, nagrań audio rozmówców symulowanego telefonu alarmowego oraz zgromadzenia metadanych setek nagrań.

## 4. Wiedza kandydata

Praca w ogólności odzwierciedla duży poziom wiedzy, erudycji i doświadczenia praktycznego jej Autor a w przedmiotowej dwudzielnej tematyce badań.
W rozdziale 1 został opisany obszar badań, cele, teza naukowa i zakres badań. Problemy, które zostały rozwiązane w pracy, odniesiono do stanu wiedzy. Autor zna i poprawnie przytacza metody tworzenia obrazów stereoskopowych. Cytowana bibliografia obca jest odpowiednio dobrana i obszerna, jednak niestety pominięte zostały aspekty percepcji – zmęczenia, rywalizacji siatkówkowej, przekłamania kolorów.

W dziedzinie przetwarzania metadanych Autor powołuje się na liczne publikacje dotyczące rozpoznawania osób, indeksowania nagrań oraz kilku metryk porównawczych. Zabrakło bardziej zaawansowanego, zgodnie ze stanem wiedzy, zastosowania metod statystycznych oraz metod drążenia danych – dyskretyzacji, oceny przydatności atrybutów, budowy i zastosowań sieci neuronowych. Autor bardzo zdawkowo i dopiero w rozdziale 4.4 odnosi się do publikacji, w których już rozwiązywano problemy korelacji, podobieństw, miar odległości, porównań wielokryterialnych.

## 5. Inne uwagi

Drobne uwagi merytoryczne:

1. Autostereoskopia – wyprowadzenie wzoru (2.18) zakłada "względnie mały kąt theta1". Należy pokazać dla jakich wartości theta1 równość (2.18) staje się nieprawdziwa. Strony 19-26 poświęcone są wyliczeniom dystansu i wymiarów bariery paralaksy i siatki soczewek dla ekranu autostereoskopowego, jednakże wyliczenia te nie mają praktycznego zastosowania w pracy i mogłyby być pominięte. Ponadto na str. 24 założenie "In Fig. 2.7 two triangles are similar, OKL and OMN, because every angle of the OKL triangle has the same measure as the corresponding angle in the OMN triangle" pomija fakt, że element soczewkowy jest z materiału o innej gęstości optycznej niż powietrze i nastąpi załamanie światła powodujące, że założenie przestaje być prawdziwe.
2. Wzory 2.37 i 2.38 opisują odwrotną i prostą proporcjonalność, którą autor oznacza we wszystkich wzorach słowami "related with", gubiąc informację o tym kiedy jest to odwrotna a kiedy prosta zależność. Zastosowany powinien być symbol $\propto$ lub $\sim$ oznaczający proporcjonalność a prawa strona umieszczona w mianowniku pod jedynką, lub ewentualnie lewa strona równa sie k podzielone przez prawą stronę, gdzie k to współczynnik proporcjonalności.
3. Asymetria na wykresach 2.20 nie została wyczerpująco skomentowana przez autora.
4. Dla rysunku 2.25 i jego opisu w teście autor pisze „Relation between visibility of the 3D effect and image quality", tymczasem trafniej byłoby „perceived subjective image quality", ponieważ to nie jest jakość narzucana np. w wyniku kompresji, tylko postrzegana przez widza po konwersji do 3D.
5. Zmienne $a_j$ i $b_{ij}$ są niepotrzebnie wielokrotnie definiowane w tym samym rozdziale – str. 75, 76, 77.
6. Niektóre treści są nie w pełni uporządkowane, np. arbitralną zależność prawdopodobieństw od MTT Autor podaje na str. 80 dla 6 wartości a bardziej szczegółowo wprowadza we wzorze 5.1 na str. 114, na 11 wartości.
7. Rozdział 5.1.3 na ponad 6 stronach przedstawia prostą charakterystykę zgromadzonej bazy nagrań, opartą głównie na histogramach. W całości mógłby być przeniesiony do załącznika, gdyż w tej postaci zaburza płynność czytania rozprawy.
8. Niejasna jest liczba osób dzwoniących rozpoznawanych w teście. Miejscami Autor pisze o 192 lub 166 osobach, a wykres na rys. 5.13 prezentuje ranking ponad 350 osób.

Ocena poprawności pisania w j. angielskim:

1. Praca napisana w j. angielskim, na ogół poprawnie i starannie, ale prezentowane interfejsy aplikacji na rysunkach 2.19 , 2.30, 2.37 są opisane w j. polskim.
2. Niewłaściwe stosowanie rodzajników, zwykle braki rodzajników nieokreślonych "a", "an" tam, gdzie one są wymagane. Zdarzają się niewłaściwe formy bierne (str. 9: "It is demand...")

3. Liczne niejasne zdania, np. "The strong 3D effect with well-seen values of perceived depth is the way to enhance the operator's concentration." Czym są tu "dobrze widoczne wartości postrzeganej głębi"?
4. Niegramatyczne zdania uniemożliwiają niekiedy zrozumienie myśli autora, np.:
    a. str. 10,"For example, the recognition of the set of similar calls should be based on similarities of various but as well as relevant data (even multimedia data like audio, video, and text) among conversations."
    b. "The authors proposed some rules, understood as possibility of occurrence particular feature value in case of occurrence other feature values."
    c. „High values of comparing scores, which correspond to the results of comparisons are assigned for the Best matching results." (str. 75.)
    d. "The GUI is screened in Fig. 2.30"
    e. „In the third set (set 3c, Table 5.9) the weights for the features that are still strongly related to the caller but are stated by the operator increased."
5. Odnośnik do bibliografii często niewłaściwie traktowany jako część zdania (str. 10: "mentioned in [Garc2007]").
6. Błędy literowe, np.: "to the knowledge to the author of this dissertation" powinno być "to the knowledge of the author of this dissertation"
7. Niewłaściwe używanie słowa „parameter", np. „$a_j$ is a parameter of j-th feature" – powinno być „$a_j$ is a value of j-th feature" – dosłownie jest „wartością j-tej cechy". Parametr z kolei to wielkość, która determinuje i konfiguruje sposób zadziałania algorytmu np. liczba iteracji, tempo nauki, a także cechy urządzenia, np. parametry techniczne.
8. W podpisie rys. 5.7 literówka „end" zamiast „and"

Drobne potknięcia redakcyjne:

1. Rys. 1.1 umieszczony przed cytowaniem
2. Niska czytelność wykresów 2.31 – 2.3.
3. Niekonsekwentne formatowanie, np. Wielkość czcionki w 1.3.1 inna niż w 2.8.1.
4. Nieprawidłowy odnośnik do rys. 3.8, w tekście (str. 70) jest: "The task was to count people up to the point indicated by the arrow shown in Fig. 7."
5. Wiele obrazów w dostarczonym pliku PDF i na wydruku ma zbyt małą rozdzielczość i recenzentowi nie udało się w tych przypadkach poprawnie dostrzegać efektu 3D. Wskazane byłoby udostępnić oryginały obrazów i wszystkie napisane aplikacje z instrukcjami ich użycia.

## 6. Podsumowanie

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami) moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego?

| Zdecydowanie TAK | Raczej TAK | Trudno powiedzieć | Raczej NIE | Zdecydowanie NIE |
|---|---|---|---|---|
| | X | | | |

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka lub Automatyka i Robotyka?

| Zdecydowanie TAK | Raczej TAK | Trudno powiedzieć | Raczej NIE | Zdecydowanie NIE |
|---|---|---|---|---|
| | X | | | |

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej?

| Zdecydowanie TAK | Raczej TAK | Trudno powiedzieć | Raczej NIE | Zdecydowanie NIE |
|---|---|---|---|---|
| | X | | | |

Stwierdzam, pomimo stosunkowo licznych uwag polemicznych o zróżnicowanym znaczeniu, które zwarłem w recenzji, że przedstawiona do oceny praca pana Juliana Balcerka zawiera odpowiedni ładunek merytoryczny i znaczną liczbę wartościowych wyników badawczo-eksperymentalnych, dzięki czemu spełnia wymagania stawiane rozprawom doktorskim i może być dopuszczona do publicznej obrony.