

Łódź, 29.01.2022

Prof. dr hab. inż. **Krzysztof Ślot**
Instytut Informatyki Stosowanej
Politechnika Łódzka

Recenzja rozprawy doktorskiej

Mohammadreza Azimi

Investigation into the Reliability of Contactless Biometric Systems

1. Tematyka i cele rozprawy

Przedłożona do recenzji rozprawa doktorska jest poświęcona zagadnieniom analizy wpływu wybranych czynników fizjologicznych i behawioralnych, stanowiących kontekst procedury rozpoznawania biometrycznego, na wiarygodność wyników tej procedury. Przedmiotem zainteresowania Doktoranta są procesy i zachowania, które mogą według Niego zniekształcać specyficzną dla każdej osoby postać danych biometrycznych uzyskiwanych na podstawie obrazów tęczówki, głosu i obrazów twarzy, a które nie były dotychczas uznawane za istotne. Ponieważ wymienione modalności są najpowszechniej stosowanymi źródłami informacji dla przeprowadzania weryfikacji lub identyfikacji biometrycznej, uwzględnienie aspektów analizy mogących skutkować dodatkowymi błędami ma kluczowe znaczenie z punktu widzenia praktyki stosowania technologii biometrycznych. Wyjaśnienie problemu istnienia negatywnego wpływu rozważonych przez Doktoranta czynników:

- zmian struktury tęczówki, które mogą być indukowane przez stany chorobowe (konkretnie – cukrzycę typu II) lub wynikają z długotrwałego wpływu palenia papierosów
- występowania zmian brzmienia głosu powiązanego z dobowym cyklem fizjologii układu oddechowego
- wzajemnego wzmacniania efektów emocjonalnej ekspresji twarzy i użycia makijażu

ma istotne znaczenie dla dziedziny biometrycznej analizy danych, dlatego zakres tematyczny rozprawy uważam za **aktualny i ważny**.

Drugim celem prac Doktoranta, stanowiącym naturalną konsekwencję pozytywnej weryfikacji postawionych hipotez, jest próba sformułowania metod redukujących negatywny wpływ wymienionych czynników. Opracowanie sposobów przeciwdziałania pogorszeniu właściwości powszechnie dzisiaj stosowanych algorytmów chroniących przed nieuprawnionym dostępem do zasobów lub pozwalających na niezawodną identyfikację osób, ma kluczowe znaczenie dla bezpieczeństwa i efektywności funkcjonowania społeczeństwa. Dlatego też uzyskanie znaczących efektów prac z pewnością stanowiłoby osiągnięcie o randze spełniającej kryteria merytoryczne stawiane przed rozprawą doktorską, będące zauważalnym wkładem do dyscypliny naukowej **informatyka techniczna i telekomunikacja**, w której bez wątpienia lokuje się obszar badań podjętych przez Doktoranta.

2. Charakter i struktura rozprawy

Przedstawiona rozprawa ma charakter eksperymentalny: Doktorant formułuje kolejne hipotezy badawcze, które następnie weryfikuje w serii przeprowadzanych doświadczeń. Z uwagi na specyficzny, unikatowy charakter stawianych tez, w większości przypadków nie istnieją repozytoria danych, mogących służyć jako podstawa testów, dlatego też istotnym i cennym elementem prac Doktoranta jest zgromadzenie i odpowiednie przygotowanie materiału doświadczalnego. Opracowane przez Doktoranta bazy danych zawierające:

- obrazy tęczy osób dotkniętych cukrzycą (723 obrazy 163 oczu), uzupełniona 1183 obrazami 360 zdrowych oczu, pozyskanych z użyciem identycznego sprzętu, stanowiącymi materiał odniesienia
- nagrania głosu osób dokonane w dwóch różnych okresach: po odpoczynku nocnym (916 próbek pobranych od 30 osób) i w okresie normalnej dziennej aktywności (864 próbki pobrane od tych samych 30 osób)
- obrazy tęczy osób palących (284 obrazy pobrane od 41 osób) i niepalących (390 obrazów, 87 osób)

stanowią niezwykle wartościowe efekty przeprowadzonych przez Niego prac, stanowiąc publicznie dostępny materiał dla badań w rozważanych kierunkach.

W odniesieniu do części eksperymentów, Doktorant skorzystał z dostępnych, adekwatnych zasobów: dla potrzeb oceny wpływu ekspresji emocji na poprawność rozpoznawania twarzy użył baz RaFD, PICS i RAVDESS, a w eksperymentach oceny możliwości rozpoznawania twarzy w obliczu zmian wywołanych starzeniem użył fragmentu bazy CelebA.

Metodyka przeprowadzania eksperymentów jest poprawna i dobrze udokumentowana, co daje możliwość powtórzenia przeprowadzonych doświadczeń. Doktorant szczegółowo informuje o warunkach procesu akwizycji danych oraz o strukturze rejestrowanego materiału. Do realizacji zadań przygotowania 'surowych' danych do analizy biometrycznej Doktorant stosuje właściwe metody przetwarzania obrazów lub sygnałów i korzysta ze sprawdzonych implementacji odpowiednich algorytmów. Wreszcie, wykorzystane metody analizy danych i opracowania wyników – algorytmy klasyfikacji oraz narzędzia statystycznej obróbki, wraz z ich implementacjami programowymi, są rozwiązaniami uznawanymi za najskuteczniejsze z punktu widzenia osiągnięcia postawionych celów.

Ogólna struktura pracy jest zgodna z oczekiwaniami – po wprowadzeniu, identyfikującym kontekst prac, definiuje tezy pracy, a następnie, w kolejnych rozdziałach, szczegółowo omawia badania przeprowadzone w celu ich wykazania. Ponieważ Autor zajmuje się trzema wyraźnie różniącymi się obszarami tematycznymi, wydaje się, że bardziej naturalną formą uporządkowania pracy byłby podział na trzy części, każda z nich dotycząca innej rozważanej modalności biometrycznej, ale zdaję sobie sprawę, że jest to uwaga o charakterze raczej polemicznym (Autor poświęca trzy osobne rozdziały na omówienie różnych wątków związanych z dokonywanymi przez siebie analizami tęczy).

Dla prezentacji wyników prac dotyczących kolejnych wątków podejmowanych w rozprawie, Doktorant przyjmuje formułę cyklu omówień treści, jakie zawarł w różnych opublikowanych przez siebie artykułach, którą uważam za niezbyt fortunną. Moje główne zastrzeżenie do takiego, dopuszczalnego w ogólności układu, dotyczy braku niezbędnego dla takiego zamysłu uporządkowania materiału, skutkującego wielokrotnym powielaniem tych samych treści w różnych miejscach pracy. Pojawiające się w każdym z rozdziałów poświęconych kolejnym wątkom rozprawy podrozdziały wprowadzające konkurują zawartością, zarówno ze sobą, jak i z zawartością Rozdziału 2, zawierającego przegląd kontekstu prac Doktoranta.

Oprócz omówienia trzech wątków badawczych, odpowiadających sformułowanym przez Doktoranta tezom, zdecydował się on dodatkowo na umieszczenie w pracy również podsumowania przeprowadzonych przez siebie badań w trzech innych obszarach, niepowiązanych z tezami rozprawy lub zakończonych brakiem konstruktywnych wniosków. Materiał ten został umieszczony w trzech załącznikach do pracy, po rozdziale zawierającym podsumowanie.

3. Tezy rozprawy

Doktorant formułuje w rozprawie trzy tezy, które w ostrożny sposób identyfikuje jako ‘sub-statements’, i które odpowiadają trzem wymienionym wcześniej obszarom badań:

S1. “Iris recognition is less effective for people with diabetes type II, regardless of their gender and age. By classifying the illness-affected samples, the performance of the system will improve.”

S2. “Morning voice can challenge the reliability of text independent speaker recognition systems. We can reduce this effect by detecting morning voice.”

S3. “The simultaneous effects of makeup and facial expressions can affect the matching accuracy of a facial recognition system. Reliability can be enhanced using makeup detectors”

Sens przedstawionych tez jest jasny w odniesieniu do ich pierwszych części, gdzie Doktorant orzeka kolejno o negatywnym wpływie cukrzycy na poprawność analizy biometrycznej tęczówki, upośledzenia poprawności rozpoznawania głosu dla próbek pobieranych rano (po okresie snu) i pogorszeniu poprawności wyniku rozpoznawania w wyniku jednoczesnego wystąpienia ekspresji twarzy i makijażu. Drugie części tez w odniesieniu do S2 i S3 są raczej skrótem myślowym: poprawa skuteczności rozpoznawania może nastąpić w wyniku odpowiedniej akcji następującej po detekcji wystąpienia wskazanego przez Doktoranta czynnika, a nie w wyniku samej detekcji. Niestety, druga część pierwszej tezy jest niejasna i nie do końca rozumiem co Doktorant na myśli pisząc, że „klasyfikując próbki dotknięte chorobą poprawimy jakość działania systemu” (tutaj skrót myślowy jest chyba bardziej skomplikowany).

Pomijając sformułowane powyżej zastrzeżenie, chcę podkreślić **oryginalność** podjętych problemów oraz **zasadność** podjęcia przedstawionych w tezach tematów, bowiem stanowią one prawdopodobne konteksty przeprowadzania analizy biometrycznej, których ewentualny wpływ na jej wyniki był do tej pory całkowicie pomijany. Wykazanie istnienia takiego wpływu i propozycja metod redukcji wynikających z tego negatywnych konsekwencji może mieć istotne znaczenie dla praktyki stosowania istniejących systemów biometrii.

Oprócz przedstawionych tez rozprawy, Autor identyfikuje również trzy dodatkowe, inne cele swoich prac: sprawdzenie hipotezy o wpływie palenia papierosów na poprawność rozpoznawania tęczówki, identyfikację emocjonalnych wyrazów twarzy o najsilniejszym negatywnym wpływie na poprawność rozpoznawania oraz hipotezy o możliwości poprawy identyfikacji twarzy osób zmienionych w wyniku starzenia z pomocą zastosowania metod sztucznego ‘postarzania’ wzorców biometrycznych

4. Merytoryczna ocena pracy

We wstępie oceny, chciałbym przez chwilę odnieść się do formy przekazu treści. Ponieważ praca napisana jest w języku angielskim, który nie jest językiem ojczystym Doktoranta, treści rozprawy nie zawsze są przekazywane w sposób dostatecznie jasny, a używana nomenklatura budzi czasami wątpliwości. Przykładowo, Doktorant identyfikuje rozważane przez siebie czynniki jako należące do ogólniejszej kategorii ‘społecznych’ (social factors), co w mojej opinii nie ma uzasadnienia.

Kwalifikacja rozważanych w pracach Doktoranta aspektów analizy biometrycznej: schorzeń, fizjologicznych powodów zmienności brzmienia głosu, czy też starzenia do kategorii czynników ‘społecznych’ wynika prawdopodobnie z innego sposobu rozumienia przez Niego tego terminu, a podjęta próba ‘systematyzacji’ obszaru podjętych badań jest w mojej opinii zupełnie niepotrzebna.

Wspólnym mianownikiem przeprowadzonych przez Doktoranta prac jest przyjęcie jednakowego schematu postępowania w analizie kolejnych podejmowanych problemów: najpierw dokonuje przeglądu stanu wiedzy, następnie stawia hipotezę odnośnie negatywnego wpływu pewnego czynnika na poprawność rozpoznawania biometrycznego i dokonuje jej eksperymentalnej weryfikacji, wreszcie, próbuje przedstawić pomysły na zredukowanie stwierdzonych, negatywnych skutków.

Stanowiący punkt wyjścia dla prezentacji prac Doktoranta przegląd stanu wiedzy w rozważanej tematyce jest bardzo obszerny, aktualny i kompetentny (choć, jak wcześniej wspomniałem, rozproszony na przestrzeni całej pracy). Niewątpliwą zaletą przeglądu jest przytoczenie szeregu mniej znanych, a bardzo zasadnych dla rozprawy wątków, takich jak występujące u osób starszych problemy z obsługą nowoczesnych rozwiązań technologicznych, skutkujące na przykład niewłaściwą prezentacją biometryk (linii papilarnych, tęczyówek) systemom akwizycji lub problem zanikającej kompatybilności stosowanych powszechnie metod wstępnego przetwarzania danych wywołany rozwojem technologicznym i pojawianiem się sensorów o innych właściwościach (tzw. ‘sensor aging’). Obydwa te zagadnienia są ściśle powiązane ze specyfiką rozpoznawania biometrycznego z użyciem urządzeń mobilnych, której dotyczy przeważająca część rozważań rozprawy. Jedyną pułapką, w którą w mojej ocenie wpada Doktorant dokonując przeglądu stanu wiedzy, jest wprowadzenie sztucznego rozróżnienia między pojęciami ‘starzenia się cech’ (‘physiological aging’) i ‘starzenia się wzorców biometrycznych’ (‘template aging’), które nie ma żadnego uzasadnienia. Źródłem zamieszania jest używanie w literaturze różnych terminów, sugerujące istnienie dwóch niezależnych procesów związanych z upływem czasu i negatywnie wpływających na rozpoznawanie. Tymczasem termin ‘starzenie się wzorca biometrycznego’ stanowi w istocie skrót myślowy, wyrażający obniżenie reprezentatywności wzorca biometrycznego pobranego w pewnej chwili czasu, wynikające ze zmian następujących w odniesieniu do danej cechy biometrycznej w czasie (długo- lub krótkoterminowych), np. powodowanych starzeniem się tkanek i narządów zmieniających ich strukturę, lub wynikającym z tego zmianom behawioralnym i zmianom stylu wykonywania czynności (np. pisma).

Metodyka przeprowadzania eksperymentów w zakresie fazy przetworzenia pozyskiwanych danych do postaci wzorców biometrycznych, podlegających klasyfikacji, jest opisana w sposób zdawkowy. Wynika to z faktu używania przez Doktoranta gotowych implementacji różnych algorytmów, co sprawia, że zamieszczane we wszystkich rozdziałach wyjaśnienia używanych metod są powierzchowne (zwykle są to wycinkowe cytaty ze źródeł) i nie mają waloru poznawczego.

4.1 Ocena wkładu w zakresie wpływu cukrzycy na poprawność rozpoznawania tęczyówki

Rozważania dotyczące pierwszej z tez rozprawy: „Iris recognition is less effective for people with diabetes type II, regardless of their gender and age”, zostały umieszczone w trzech kolejnych rozdziałach pracy (3,4 i 5), przy czym rozdział 3 jest poświęcony eksperymentom porównania wyników analizy biometrycznej osób zdrowych i osób dotkniętych schorzeniem, rozdział 4 rozszerza badania o analizę wpływu wieku na wyniki rozpoznawania, zaś rozdział 5 podejmuje dodatkową analizę wpływu płci.

Przeprowadzone przez Doktoranta eksperymenty, opisane w Rozdziale 3, polegały na przeprowadzeniu analizy porównawczej wyników rozpoznawania tęczyówki w obrębie dwóch różnych grup: dla osób zdrowych i dla osób dotkniętych chorobą. Uzyskane wyniki Doktorant poddał poprawnej metodyce statystycznej oceny, sprawdzając zasadność dwóch hipotez

‘zerowych’: o tym, że w obrębie obydwu rozważanych grup średnia dokładność rozpoznawania jest taka sama, a następnie o tym, że wyniki dla obydwu grup pochodzą z tego samego rozkładu. W pierwszym przypadku zastosował do weryfikacji test t-studenta, w drugim – test Kołmogorowa-Smirnowa. Uzyskane wyniki wskazują na prawdziwość hipotezy alternatywnej, o występowaniu różnic między poprawnością analiz w odniesieniu do rozważanych grup, co nie znaczy, że potwierdzają hipotezę o przyczynie tego stanu, jakim jest występowanie schorzenia. Doktorant ma świadomość możliwej rozbieżności i bardzo trafnie wskazuje potencjalny alternatywny powód uzyskiwanej różnicy - różną strukturę wiekową osób należących do rozważanych grup: w grupie osób ze schorzeniem znajduje się więcej osób starszych niż w grupie osób młodych. Decydując się na wyjaśnienie wpływu wieku na rozpoznawanie, przeprowadził kolejne eksperymenty, które wykazały słuszność innej hipotezy, o statystycznie istotnej wrażliwości wyników rozpoznawania tęczówki w odniesieniu do wieku osób. Wykazana przez Doktoranta wnikliwość badawcza – nie zaakceptował pozornie rozstrzygających wyników eksperymentów, jest godna uznania.

Paradoksalnie, wnikliwość ta, stanowiąca największą w mojej opinii wartość treści przedstawionych w rozdziałach dotyczących biometrycznej analizy tęczówki, ujawnia też największą słabość rozważań – brak eksperymentów, konfrontujących wyniki analizy właściwie dobranego materiału eksperymentalnego. Aby uzasadnić brak dostatecznej podstawy do potwierdzenia postawionej tezy, trzeba usystematyzować zakres przeprowadzonych przez Doktoranta eksperymentów. Otóż, dzieli On zbiór posiadanych przykładów na cztery części: pierwsza dychoomia dokonana jest według kryterium zdrowia, skutkując dwoma rozłącznymi podzbiorami obrazów oczu osób zdrowych (Z) i chorych (C), które następnie są dalej dzielone na dwie kategorie wiekowe – zdrowi młodzi (ZM), zdrowi starzy (ZS) oraz analogicznie, chorzy młodzi (CM) i chorzy starzy (CS). Korzystając z utworzonych podzbiorów przeprowadza następnie eksperymenty rozpoznawania biometrycznego, przy czym w rozdziałach 3 i 4 przedstawia:

1. Porównanie wyników weryfikacji biometrycznej przeprowadzonej dla dwóch doświadczeń: w pierwszym dziedzinę stanowi zbiór Zdrowi (suma zbiorów ZM i ZS), w drugim – Chorzy (suma zbiorów CM i CS).
2. Porównanie wyników dla innych dwóch przypadków: dziedzina pierwszego to ZM, dziedzina drugiego to ZS
3. Porównanie wyników dla dwóch kolejnych przypadków: dziedzina pierwszego to CM, dziedzina drugiego to CS

Na podstawie przedstawionych analiz, Doktorant wysnuwa wniosek, że poprawność analizy biometrycznej tęczówki jest zależna od występowania schorzenia – cukrzycy. W mojej opinii ten wniosek byłby uzasadniony tylko wtedy, gdyby została przeprowadzona analiza porównawcza wyników dwóch eksperymentów: rozpoznawania w obrębie grupy ZM i CM (czyli – czy istnieje statystycznie istotna różnica w rozpoznawaniu ‘młodych’ tęczówek osób zdrowych i chorych) oraz dla analogicznej pary zbiorów ZS i CS. Brak takich analiz nie daje podstaw do potwierdzenia tezy o pogorszeniu wyników weryfikacji biometrycznej dla osób u których stwierdzono cukrzycę, bowiem zaobserwowany wynik może być efektem zarówno istnienia bliżej nieokreślonych i trudnych do powtarzalnej rejestracji struktur pojawiających się w tęczówce osób chorych, ale może być efektem nadreprezentacji osób starszych w klasie ‘chorzy’, lub wreszcie, może być wypadkową obydwu przyczyn. Ta nadreprezentacja osób starszych ma, jak zresztą wcześniej zauważa Doktorant, negatywne implikacje dla procesu rozpoznawania, wynikające np. z trudności w poprawnej prezentacji wzorców lub problemów z akomodacją żrenicy.

Chcę podkreślić, że nie twierdzę, że sformułowana teza S1 rozprawy jest nieprawdziwa, a jedynie, że Doktorant nie wykorzystał okazji do przeprowadzenia eksperymentów rozstrzygających rozważaną kwestię.

Zagadnienia przedstawione w Rozdziale 5 koncentrują się wokół analizy wpływu płci na wyniki analizy biometrycznej tęczówki. Doktorant przedstawia w niej wyniki analiz, przeprowadzanych

dla osób zdrowych i dotkniętych schorzeniem z podziałem na płeć. Przykładowo, pierwszy zbiór wyników to ocena podobieństw rozkładów uzyskanych w eksperymencie „Women genuine”, gdzie w Tabeli 5.1 Doktorant podaje parametry statystyczne (wartości średnie i odchylenia) dla rozkładów uzyskanych dla porównań tęczywek zgodnych (genuine) u kobiet zdrowych i kobiet chorych. Parametry statystyczne wskazują faktycznie na ‘lepsze’ wartości uzyskane dla kobiet zdrowych (mniejsza wartość średnia i mniejsza wariancja), a ich statystyczna istotność jest potwierdzona przeprowadzeniem odpowiednich testów (t-Studenta i Kołmogorowa-Smirnowa). Takie same wnioski dotyczą porównania poprawności rozpoznawania zgodnych tęczywek w obrębie grupy mężczyźni oraz porównań niezgodnych tęczywek w obrębie płci dla grup ‘zdrowi’ i ‘chorzy’ (tym razem ‘lepsze’ są większe wartości średnie). Przedstawione wyniki uprawniają do stwierdzenia o niezależnym od płci pogorszeniu wyników rozpoznawania, uzyskiwanym dla osób z cukrzycą typu II. Niestety, podobnie jak poprzednio, nie można udzielić odpowiedzi, czy powodem pogorszenia poprawności rozpoznawania jest występowanie schorzenia, czy np. różna struktura wiekowa w obrębie grup ‘zdrowi’ i ‘chorzy’. Sposób podsumowania wyników eksperymentów Rozdziału 5 budzi nieco zamieszania w obliczu brzmienia tytułu („Gender-dependency of the Diabetes Effects ...”), który sugeruje poszukiwanie różnic w poprawności rozpoznawania tęczywek osób chorych w funkcji płci. Aby udzielić odpowiedzi na tak zadane pytanie, należałoby porównać odległości między rozkładami wyników dopasowań tęczywek zgodnych i różnych (genuine vs. Impostor) dla chorych kobiet i mężczyzn. Takiego wyniku Doktorant nie przedstawia, a jego konkluzje odnoszą się do tezy sformułowanej w inny sposób w tekście rozdziału.

W drugiej części pierwszej tezy rozprawy Doktorant sugeruje możliwość zwiększenia poprawności rozpoznawania poprzez zastosowanie (jak można podejrzewać) dwuetapowej procedury analizy, w której faza rozpoznawania jest poprzedzona fazą detekcji występowania schorzenia. Niestety, próba opracowania algorytmu detekcji kończy się fiaskiem, zaś Doktorant nie próbuje udzielić żadnej odpowiedzi na pytanie o ewentualne pomysły na sposób, w jaki wykorzystałby uzyskaną wiedzę dla poprawy rozpoznawania. W efekcie, stwierdzenie „By classifying the illness-affected samples, the performance of the system will improve.” pozostaje gołosłowne.

4.2 Ocena wkładu w zakresie wpływu dobowych zmian brzmienia głosu na rozpoznawanie mówcy

Rozważania dotyczące zagadnień odnoszących się do drugiej ze sformułowanych tez rozprawy (S2. “Morning voice can challenge the reliability of text independent speaker recognition systems. We can reduce this effect by detecting morning voice.”) są przeprowadzone w podobnym schemacie jak to miało miejsce w odniesieniu do tezy dotyczącej rozpoznawania tęczywki. Doktorant prezentuje kontekst, materiał eksperymentalny, narzędzia programowe używane w przetwarzaniu danych i klasyfikacji, przebieg eksperymentów i sposób analizy wyników. Tym razem jednak, metodyka zastosowana do rozstrzygnięcia zasadności postawionej tezy nie budzi żadnych zastrzeżeń, a wyniki analiz są przekonujące. Wpływ zasugerowanego przez Doktoranta czynnika na poprawność weryfikacji mówcy jest, dla rozważanego materiału eksperymentalnego ewidentny. Zachowując ostrożność wynikającą z braku reprezentatywności próbki wobec populacji, można stwierdzić, że uzyskane wyniki uprawdopodobniają słuszność hipotezy S2. Ta konkluzja dowodzi doskonałej intuicji, a jednocześnie ma istotne implikacje z punktu widzenia możliwości optymalizacji pracy systemów uwierzytelniania bazującego na głosie.

W przeciwieństwie do nieudanej próby zbudowania detektora stanu chorobowego tęczywki, Doktorant z powodzeniem realizuje zasygnalizowane w drugiej części rozważanej tezy zadanie detekcji ‘porannego’ głosu. Niestety, nie wiadomo nic o szczegółach tajemniczego algorytmu, który został przez Niego wykorzystany (jedyna informacja to skromne stwierdzenie, że jest on wyrafinowany). Możliwość określenia kategorii, do której należy próbka daje oczywiście możliwość odpowiedniej reakcji i użycia dla celów weryfikacji odpowiedniego modelu mówcy (jeśli wcześniej zadbano o stworzenie odpowiednich dla pory dnia wariantów wzorców). Szkoda że

taka konkluzja nie została wyrażona przez Autora w sposób jawny, chociaż być może uznał ją za trywialną i dlatego nie rozwinął wątku sposobu wykorzystania wyników opracowanego przez siebie algorytmu detekcji.

Podobnie jak w poprzednich częściach, wprowadzenie do zasadniczych treści rozdziału jest dość pobieżne. Dodatkowo, w części 6.6 rozdziału Doktorant wskazuje cztery właściwości sygnału mowy, które według Niego istotnie wpływają na różnicowanie mowy 'porannej' i 'popołudniowej'. Jednakże pomimo uznania, że właściwości te są 'kamieniem węgielnym' (cornerstone) oceny głosu, nie widzę ich bezpośredniego wykorzystania w budowie reprezentacji głosu (argumentem analizy jest 'surowy' sygnał mowy), więc trudno powiedzieć, jaki w zamyśle Doktoranta był powód ich uwypuklenia.

4.3 Ocena wkładu w zakresie analizy łącznego wpływu makijażu i ekspresji na rozpoznawanie twarzy

W ostatnim wątku swoich badań, Doktorant weryfikuje kolejną ciekawą i nieoczywistą hipotezę o wzajemnym 'wzmacnianiu' się negatywnego wpływu na poprawność analizy biometrycznej twarzy dwóch czynników, rozważanych wcześniej niezależnie od siebie – zmian wyglądu wywołanych emocjami mimiką twarzy i makijażem (Doktorant dla stanów emocjonalnych używa niezbyt zręcznego terminu 'mood', czyli nastrój).

Dla weryfikacji zaprezentowanej tezy Doktorant, tak jak poprzednio, korzysta z gotowych narzędzi programowych. W obliczu braku odpowiedniego materiału eksperymentalnego, w celu uzyskania baz danych zawierających przykłady zawierające obydwa czynniki zaburzające wygląd, Doktorant stosuje dwie pomysłowe strategie dodawania makijażu do obrazów twarzy emocjonalnych, przy czym pierwsza wymaga manualnego nanoszenia makijażu, zaś druga realizuje to zadanie automatycznie. Jako główne narzędzie analizy Doktorant wykorzystuje wytrenowany, głęboki klasyfikator neuronowy dostępny w bibliotece dlib, przy czym jako wynik przetwarzania w sieci wykorzystywana jest wygenerowana przez którąś z warstw sieci (nie znalazłem dokładnej informacji która to warstwa) reprezentacja obrazu wejściowego. Dodatkowo, dla celów weryfikacji wyników, używa również dwóch innych gotowych klasyfikatorów. Uzyskiwane rozkłady deskryptorów stawały się następnie argumentem analiz i testów statystycznych. W ramach weryfikacji postawionej hipotezy, Doktorant poszukiwał odpowiedzi na trzy szczegółowe pytania:

- Czy jednoczesne występowanie ekspresji emocjonalnej i silnego makijażu może w statystycznie istotny sposób zaburzyć wyniki rozpoznawania twarzy?
- Która kategoria emocji w powiązaniu z obecnością silnego makijażu powoduje największe różnice względem twarzy o neutralnym wyglądzie?
- Która kategoria emocji (bez obecności makijażu) wywołuje największe odstępstwa od reprezentacji wyznaczonych dla twarzy o neutralnym wyglądzie?

Wyniki analiz uzyskane dla pierwszego wątku zobrazowane w postaci krzywych ROC na rys. 7.3 i 7.4, udzielają przekonującej odpowiedzi twierdzącej na pierwsze pytanie. Podsumowanie rezultatów eksperymentów odnoszących się do drugiego wątku szczegółowego, przedstawione na rys. 7.5 i 7.6, bazuje na analizie rozkładów wyników porównań i, jak poprzednio, jest przeprowadzone rzetelnie, z użyciem testów: Kołmogorowa-Smirnowa oraz t-Studenta. Uzyskane statystyki uprawniają Doktoranta do wysnucia poprawnego wniosku **potwierdzającego hipotezę** o statystycznie istotnym wpływie obecności makijażu na twarzy o konkretnej emocji ('radość') na wiarygodność rozpoznawania biometrycznego i braku takiego potwierdzenia w odniesieniu do innych rozważanych emocji. Lekkim zgrzytem w relacji dotyczącej omawianego wątku pracy jest poplątanie przez Doktoranta w podrozdziale 7.6 testów Kołmogorowa-Smirnowa i t-Studenta podczas identyfikacji hipotezy alternatywnej („F_withoutmakeup (Neutral vs expressive) <

F_withmakeup(Neutral vs expressive)” - ta postać hipotezy może dotyczyć testu t-Studenta, natomiast jest odniesiona do testu KS, co nie ma sensu.

Identyfikacja emocji najbardziej utrudniającej rozpoznawanie biometryczne twarzy została również przeprowadzona w sposób metodycznie poprawny i dający podstawę do sformułowania uzasadnionych wniosków. Dla każdej pary rozkładów wyników porównań deskryptorów twarzy neutralnych i deskryptorów twarzy emocjonalnych zostały wyznaczone odległości Bhattacharyya, jednoznacznie wskazując na ‘Obrzydzenie’, jako emocjonalną kategorię o reprezentacji najbardziej odległej od reprezentacji twarzy neutralnych. Interesujące jest, że ta kategoria emocjonalna nie została uznana za istotne źródło pogorszenia klasyfikacji twarzy z nałożonym makijażem.

Ostatnia część rozdziału poświęconego analizie wpływu ekspresji i makijażu na rozpoznawanie biometryczne (szkoda, że nie została wydzielona jako osobny podrozdział) dotyczy prac Doktoranta nad próbą redukcji negatywnych skutków występowania rozważanych czynników. Do rozpoznawania twarzy z makijażem Doktorant proponuje używać klasyfikatorów trenowanych na obrazach twarzy z automatycznie nałożonym makijażem. Ponieważ wybór klasyfikatora (albo trenowany na twarzach bez, albo trenowany na twarzach z makijażem) wymaga detekcji obecności makijażu, Doktorant z powodzeniem opracowuje odpowiedni detektor, tym samym spełniając obietnicę sformułowaną w drugiej części trzeciej tezy.

Podsumowanie całości pracy jest bardzo płytkie. Zamiast syntetycznej prezentacji istoty swoich dokonań i przemyśleń dotyczących np. ograniczeń zaproponowanych rozwiązań, Doktorant kopiuje w nim akapity z wcześniejszych rozdziałów pracy oraz chwali się wyrefinowaniem jednego, absolutnie szablonowego algorytmu klasyfikacji. Co więcej, treść podpunktu o nazwie „Future works”, zaczynająca się pomyłką numeracją rozdziałów, opowiada o tym co zrobił w zakresie detekcji różnic głosu i wskazuje, że gdyby dostępna była obszerniejsza baza danych, to wyniki może byłyby lepsze.

4.4 Ocena dodatkowych prac Autora opisanych w załącznikach

Po zakończeniu głównej części rozprawy, obejmującej treści które Doktorant uznał za istotę swojego osiągnięcia naukowego, w trzech kolejnych załącznikach przedstawione zostały dodatkowo zwięzłe raporty z realizacji trzech innych zadań biometrycznej analizy danych. Pierwszy z nich to próba stwierdzenia, czy palenie papierosów jest czynnikiem obniżającym poprawność rozpoznawania tęczy. Problem jest na pewno ciekawy – regularne ‘zatrucie’ organizmu być może powoduje uszkodzenia tęczy (tak jak i innych tkanek). Doktorant poprzedza prezentację swoich eksperymentów przeglądem literatury, z której wynika, że zidentyfikowanym obszarem możliwego wpływu palenia na oko są zaburzenia akomodacji (przekładające się na proces ekstrakcji obszaru tęczy). Jak to miało miejsce we wcześniejszych pracach, na uznanie zasługuje wysiłek Doktoranta włożony w opracowanie unikatowego zbioru danych eksperymentalnych. Uzyskane przez Doktoranta wyniki wskazują na brak korelacji między wiarygodnością analizy tęczy a kategorią testowanych osób (palący – niepalący).

Kolejne podjęte przez Doktoranta zadanie to próba oszacowania wpływu emocji na rozpoznawanie mówcy. Wykorzystując bazę danych nagrań odgrywanej mowy emocjonalnej, Doktorant kolejno dokonuje analiz rozkładów reprezentacji głosu, wyznaczonych przez sieć neuronową. Cele szczegółowe przeprowadzonych eksperymentów to identyfikacja emocji dających najlepsze i najgorsze ‘wewnątrzklasowe’ wyniki analizy biometrycznej (ta sama emocja jest używana do treningu i testowania), identyfikacja emocji dających najlepsze i najgorsze wyniki analizy ‘międzyklasowej’. oraz wpływu intensywności ekspresji emocji na rozpoznawanie. Dla używanej w eksperymentach bazy danych Doktorant wskazuje odpowiednie pary i potwierdza oczekiwaną zależność wiarygodności rozpoznawania od stopnia intensywności emocji. Interesującym dodatkowym wątkiem prac jest przeprowadzenie analizy sprawdzającej, czy wyznaczone reprezentacje mówców grupują się w funkcji emocji w wyraźne klastry. Potwierdzenie tego faktu

jest wartościowym wynikiem prac Doktoranta, wskazującym na pogorszenie wyników analizy biometrycznej (czynnik zróżnicowanej emocji powinien być marginalizowany przez poprawnie wytrenowany klasyfikator).

Ostatni wątek dodatkowych prac Doktoranta dotyczył weryfikacji hipotezy o możliwości poprawy rozpoznawania biometrycznego twarzy w odniesieniu do zdjęć wykonanych na przestrzeni długiego przedziału czasu, poprzez wykorzystanie oprogramowania do 'postarzania' wzorca biometrycznego. Koncepcja jest bardzo ciekawa i na pewno jej pozytywne potwierdzenie ma niebagatelne znaczenie dla dziedziny. Prezentacja wyników wprowadza niestety zamieszanie – w tekście Doktorant twierdzi, że Tabela C-1 przedstawia wyniki porównań par kategorii: młody-stary i młody-postarzony, zaś w tabeli używa par: młody-stary i postarzony-stary. W efekcie, trudno odnieść się do przeprowadzonej przez Doktoranta dyskusji wyników, natomiast bez zastrzeżeń można przychylić się do sformułowanego na końcu wniosku o zafałszowaniu biometrycznej informacji przez narzędzie do postarzania, opracowane wyłącznie pod kątem zapewnienia subiektywnie atrakcyjnego rezultatu.

5. Ocena formalnej strony rozprawy

Strona formalna przedstawionej rozprawy jest, w mojej ocenie, słaba. Doktorant ma wyraźne problemy z uporządkowaniem treści, które chce przekazać. Przyjęty sposób prezentacji zagadnień odnoszących się do tez rozprawy – wzorowanie się na tekstach opublikowanych w artykułach, powoduje wielokrotne powtórzenia oraz umieszczanie materiału dotyczącego przeglądu stanu wiedzy w różnych miejscach, nie tylko w służącym do tego celu rozdziale 'Related work'. Przykładowo, fragmenty podpunktów 3.4.1 4.4.1 są identyczne, podobnie jak 3.4.2c i 4.4.4, powtórzone są akapity dotyczące zmian głosu w p.6.2 i w Podsumowaniu, podobnie ten sam tekst zawarto w rozdziale 7.2 i Podsumowaniu. Identyczne fragmenty o rozszerzonym zbiorze danych tęczówek są podawane w części 5.1 i 5.4. Tym, którzy jeszcze nie zapamiętali, Doktorant kilkakrotnie przypomina szczegółowo identyczne informacje o sprzęcie do akwizycji obrazów tęczówki, identyczne informacje o bazie próbek mowy znajdują się na stronach 71 i 74 oraz w Podsumowaniu itp.. Testem na spostrzegawczość Czytelnika jest zapewne trzykrotne powtórzenie tej samej, równie doniosłej co niezrozumiałej myśli „Today's smartphones are equipped with biometric tech such as voice.” Powtórzenia tych samych treści zdarzają się również w obrębie jednego podrozdziału (np. 6.2). Irytuje powtarzanie we wstępie do kolejnych rozdziałów dokładnie tych samych ogólnych rozważań dotyczących wrażliwości jakości analiz biometrycznych na rozmaite czynniki i procesy środowiskowe, fizjologiczne i behawioralne (3.1, 4.1, 5.1).

Tekst bywa często chaotyczny – w niektórych miejscach pojawiają się zdania, które prawdopodobnie są pozostałością wcześniejszych wersji, bo nie mają związku z poprzednikiem i następnikiem – taki chaos mamy np. w akapicie (str. 22): „The performance of online-signature and offline handwriting recognition systems are highly dependent on the age of the user” - (zdanie dotyczy podpisu) - „This is because the physiological age can affect the face.” (następne zdanie - dotyczy twarzy). „According to a previously published paper [27], older users displayed a diminished magnitude of pen dynamics ...” (kolejne - znowu jest o podpisie). Wreszcie, w pracy znajdują się 'osobliwości' - akapity zupełnie nie powiązane z wcześniejszą ani późniejszą narracją (np. akapit o szumie i planie jego redukcji na stronie 97), a zupełnym nieporozumieniem jest ponowne wyjaśnianie Czytelnikowi szeregu wcześniej przedstawianych kwestii: czym jest tęczówka, jakie narzędzia były użyte w Podsumowaniu pracy.

W pracy jest również bardzo wiele błędów edytorskich, niezrozumiałych stwierdzeń lub tez niczym nie popartych – poniżej przedstawiam listę wyrywkowych uwag do konkretnych stwierdzeń pracy.

p.11 - The only problem with biometric solutions is their lack of performance. - ? o co chodzi Autorowi?

p.12 - Age has a strong psychological basis ... - nie rozumiem

p.26: „However, due to the different photonic structure, ... ” phonemic?

p.37: „... four samples from each group have been presented and as it can be illustrated the samples captured using near infra-red camera.” - nie rozumiem

38: „Due to Differences of Discrete Cosine Transform’s much lower complexity, it can be considered as a computationally intensive replacement for the Karhunen Loeve Transform”

Wygląda to na błąd logiczny - ‘z uwagi na dużą mniejszą złożoność jest złożonym obliczeniowo zamiennikiem ...’

38: „it calculates the truncated Chebyshev series processing minimax properties.” - nie rozumiem tego wyjaśnienia, poświęconego wynikowi analizy DCT.

39: Kolmogorov Simonov – powinno być Kolmogorov-Smirnov

48: Wyjaśnienie treści rys.4.1 w tekście jest sprzeczne z informacją w podpisie rysunku

73: Today’s smartphones are equipped with biometric tech such as voice. - nie rozumiem co Autor ma na myśli (to samo zdanie występuje również we wcześniejszej części pracy).

73. The only problem with this biometric solution is its lack of performance. - identyczne jak na str.11, niezrozumiałe

75: „... which is a powerful speaker recognition deep network, using ‘thinResNet’ trunk architecture and a dictionary-based NetVLAD or GhostVLAD layer to aggregate features across time, which can be trained end-to-end. - całe zdanie jest skopiowane z Abstraktu artykułu [169] (odwołanie do tego artykułu znajduje się w innym, niezbyt właściwym miejscu).

76: Brak wyjaśnienia znaczenia symboli w równaniu 6.3 – wydaje się, że dla oznaczenia funkcji autokorelacji użyto symbolu AC_V . Nie jest również jasne, jaki okres T ma na myśli Doktorant – czy jest to wartość średnia okresów częstotliwości podstawowej? Czy może jakiś inny parametr?

77: Nie rozumiem opisów:

- „In order to obtain the mean pitch (the fundamental frequency) of each sample, 10 seconds was split;” - Autor wspominał o 6-sekundowych nagraniach, więc skąd się biorą fragmenty 10-sekundowe, które są ‘dzielone’?

- „... the speech activity was detected first – by determining the pitch range between 75 to 500 hertz and using autocorrelation analysis method – then for each 10 ms of sample the F0 was obtained.” Jak można wykryć mowę ‘ustalając zakres częstotliwości tonu podstawowego na 75-500Hz’? To jest chyba jakiś skrót myślowy. Co więcej, jeżeli analizowany odcinek trwa 10 ms, to minimalna częstotliwość jaką w nim można wykryć to 100 Hz, więc co się dzieje z częstotliwościami od 75Hz do 100Hz?

85: Autor kopiuje fragmenty tekstu z artykułu bez korekt dostosowujących treść do nowej formy przekazu (np. This paper powinno być zamienione na This part/chapter, The main contributions of this manuscript ...)

90: „The software works manually (the samples must be processed individually and it is possible for the user to change the style of full makeup), but everything can be automatically done using the dlib code.” - zdanie wydaje się być logicznie sprzeczne: po pierwsze, do końca nie rozumiem, co znaczy, że program ‘działa ręcznie’ - czy oznacza to, że wirtualny makijaż jest nakładany ręcznie? Jeżeli tak, to jak się ma do tego stwierdzenia ciąg dalszy, czyli - ‘ale wszystko może być zrobione automatycznie’? Czy Autorowi chodziło o przeciwstawienie możliwości dwóch różnych programów, które używał? To już przecież wcześniej napisał, więc nie wiem, po co jeszcze raz przedstawiać tą samą treść (i to jeszcze w średnio zrozumiałym sposób).

90: „The maximum possible number of comparison scores will be achieved between neutral facial images, and expressive facial images (the number of neutral facial images times the number of expressive facial images)” - zdanie jest zbudowane nieudolnie – Autor chyba chce powiedzieć to, co zapisał w nawiasie, czyli, określić liczbę porównań, których będzie dokonywać.

92: „In order to validate the results to more matches ...” - ‘... using other matchers ..’ ?

94: „Which emotions can make the genuine scores from the application of lipstick makeup on facial images statistically significant? - to jest skrót myślowy: na czym polega statystyczna istotność wyników porównań obrazów z makijażem dla danej emocji? Pewnie Doktorant ma na myśli ocenę, czy różnice między wynikami rozpoznawania dla twarzy bez makijażu i twarzy z makijażem są statystycznie istotne, ale to nie wynika z przytoczonego zdania.

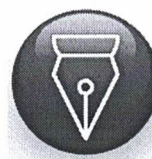
92: Co dokładnie Autor ma na myśli pisząc o ‘Frobenius normalization of vector spaces’?

...

6. Wniosek końcowy

Podsumowując przedstawioną recenzję, chcę stwierdzić, że w pracy znajdują się elementy wartościowe, mające istotne znaczenie z punktu widzenia dziedziny biometrycznej analizy danych, jak również elementy o znaczeniu drugorzędym. Doktorant w metodycznie poprawny sposób dowodzi słuszności dwóch z trzech postawionych przez siebie tez, przy czym w odniesieniu do trzeciej tezy, brak wyjaśnienia jej słuszności wynika z pogubienia się Doktoranta w mnogości przeprowadzonych przez Niego eksperymentów. Tezy stawiane przez Doktoranta są oryginalne i dotyczą bardzo powszechnych, choć dotychczas ignorowanych aspektów analiz, a wykorzystanie wniosków płynących z przedstawionych prac daje szansę na zwiększenie poprawności działania powszechnie stosowanych systemów rozpoznawania biometrycznego. Efektem prac Doktoranta jest również opracowany przez Niego unikatowy materiał eksperymentalny dla dalszych badań w dziedzinie (siłą rzeczy jest on dość ograniczony, co sprawia, że w formułowaniu wniosków końcowych czasami wskazana byłaby większa wstrzeźliwość).

Konkludując recenzję, chciałbym stwierdzić, że przedłożona do recenzji rozprawa doktorska Pana Mohammadreza Azimiego pt. „Investigation into the Reliability of Contactless Biometric Systems” **spełnia** w moim przekonaniu wymagania określone w odnośnej ustawie o stopniach i tytule naukowym i tym samym **wnioskuje o dopuszczenie Autora rozprawy do publicznej obrony**.



Signed by /
Podpisano przez:

Krzysztof Maciej
Ślót
Politechnika Łódzka

Date / Data: 2022-
01-29 14:13

Krzysztof Ślót