

Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych

Warszawa, 23 października 2018 r.

D z i e k a n a t

Uprzejmie informuję, że na Wydziale Elektroniki i Technik Informacyjnych Politechniki Warszawskiej odbędzie się w dniu 6 listopada 2018 r. publiczna obrona rozprawy doktorskiej

Mgr. inż. Marka Kowalskiego

temat: „Localization and tracking of facial landmarks in images and video sequences”.

promotor – prof. dr hab. inż. Władysław Skarbek z Politechniki Warszawskiej

recenzenci:

prof. dr hab. inż. Bogusław Cyganek z Wydziału Informatyki, Elektroniki i Telekomunikacji
Akademii Górniczo – Hutniczej

dr hab. inż. Paweł Forczmański z Wydziału Informatyki Zachodniopomorskiego
Uniwersytetu Technologicznego w Szczecinie

Obrona odbędzie się w dniu 6 listopada 2018 r. w sali 116 na Wydziale Elektroniki i Technik Informacyjnych – Gmach im. Janusza Groszkowskiego, Warszawa, ul. Nowowiejska 15/19; początek godz. 10³⁰.

Po adresem: www.elka.pw.edu.pl/Wydzial/Rada-Wydzialu/Harmonogram-obron-doktorskich-streszczenia-i-recenzje zapewniony jest na stronie Wydziału dostęp do tekstów streszczenia rozprawy i recenzji, jak również do tekstu rozprawy umieszczonej w Bazie Wiedzy Politechniki Warszawskiej.

Dziekan



prof. dr hab. inż. Krzysztof Zaremba

Mgr inż. Marek Kowalski

Tytuł rozprawy doktorskiej: „Localization and tracking of facial landmarks in images and video sequences”.

„Lokalizacja i śledzenie punktów charakterystycznych twarzy w obrazach i materiałach wideo”

Promotor: prof. dr hab. inż. Władysław Skarbek

Streszczenie:

Poniższa praca opisuje dwie nowe metody lokalizacji i śledzenia punktów charakterystycznych twarzy. Pierwsza metoda opiera się o nową metodę regresji nazwaną K-klastrowe Drzewa Regresyjne z Ważonym Podziałem, oraz o metodę inicjalizacji odporną na zróżnicowane warunki. Druga metoda, nazwana Głęboką Siecią Punktów Twarzy, opiera się o wieloetapową głęboką sieć neuronową. Cechą charakterystyczną obu metod jest wysoka odporność na takie utrudnienia jak: trudna inicjalizacja, nie frontalna poza głowy czy nierównomierne oświetlenie. Wysoka odporność Głębokiej Sieci Punktów Twarzy jest udowodniona przez fakt, iż metoda ta ma o ponad 70\% niższą stopę niepowodzenia na zbiorze 300-W public test set niż wcześniejsze metody. Poza szczegółowym opisem proponowanych metod praca zawiera także wyniki szerokiej gamy testów, oraz opis dwóch wybranych zastosowań stworzonych metod, które zostały zaimplementowane podczas trwania doktoratu.

Szczecin, 5 czerwca 2018r.

dr hab. inż. Paweł Forczmański
Katedra Systemów Multimedialnych
Wydział Informatyki
Zachodniopomorski Uniwersytet
Technologiczny w Szczecinie
ul. Żołnierska 52
71-220 Szczecin

RECENZJA ROZPRAWY DOKTORSKIEJ
DLA RADY WYDZIAŁU ELEKTRONIKI I TECHNIK INFORMACYJNYCH
POLITECHNIKI WARSZAWSKIEJ

Tytuł rozprawy: **Localization and tracking of facial landmarks in images and video sequences**

Autor rozprawy: **mgr inż. Marek Kowalski**

Przedmiotem niniejszej recenzji jest rozprawa doktorska zatytułowana "Localization and tracking of facial landmarks in images and video sequences", napisana w roku 2018, której autorem jest mgr inż. Marek Kowalski.

Promotorem pracy jest prof. dr hab. inż. Władysław Skarbek, a promotorem pomocniczym jest dr inż. Jacek Naruniec. Niniejsza recenzja została przygotowana na zlecenie prof. dr hab. inż. Krzysztofa Zaremby, Dziekana Wydziału Elektroniki i Technik Informacyjnych, zawarte w piśmie z dnia 6 kwietnia 2018r., w związku z decyzją Rady Wydziału z dnia 27 marca 2017r. Recenzja została przygotowana w formie odpowiadającej kwestionariuszowi przesłanemu przez Zleceniodawcę.

I. Jakie zagadnienie naukowe jest rozpatrzone w pracy /teza rozprawy/ i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?

Praca umiejscowiona jest w obszarze widzenia komputerowego, przetwarzania obrazów i uczenia maszynowego. Jej tematyka skupia się wokół dwóch związanych ze sobą zagadnień: algorytmów lokalizacji punktów charakterystycznych na obrazach twarzy (landmarków) i ich śledzenia. W pracy przedstawiono dwa oryginalne algorytmy. Pierwszy z nich opiera się na metodzie regresji i dopasowanej metodzie inicjalizacji odpornej na zróżnicowane warunki początkowe – w pracy występuje pod nazwą K-klastrowe Drzewa Regresyjne z Ważonym Podziałem (K-cluster Regression Forests with Weighted Splitting). Drugi algorytm bazuje na wieloetapowej głębokiej sieci neuronowej wykorzystującej, oprócz standardowych danych, mapę cieplną punktów charakterystycznych i mapę cech obrazu twarzy – w pracy występuje pod nazwą: Głęboka Sieć Punktów Twarzy (Deep Alignment Network). Oba algorytmy w swoich założeniach mają charakteryzować się wysoką od-



pornością na trudną inicjalizację, nie-frontalne ustawienie (pozę) głowy, znaczące zmiany w ekspresji, czy też nierównomierne oświetlenie.

W swojej pracy (Rozdział 1.3) Autor definiuje cele pracy, określone jako:

1. opracowanie i realizacja metody lokalizacji charakterystycznych punktów twarzy, która posiadać będzie niższy wskaźnik błędów niż współczesne metody state-of-the-art;
2. opracowanie i realizacja metody śledzenia charakterystycznych punktów twarzy, która posiadać będzie niższy wskaźnik błędów niż współczesne metody state-of-the-art.

W Rozdziale 1.4 zaprezentowane są natomiast powiązane z powyższymi tezy:

1. Zaproponowana metoda lokalizacji i śledzenia charakterystycznych punktów twarzy – Głęboka Sieć Punktów (Deep Alignment Network) osiąga niższy poziom błędów niż najlepsze znane obecnie metody;
2. Zastosowanie, na etapie inicjalizacji, zaproponowanej metody Trójwymiarowej Afinicznej Regresji Pozy (3D Affine Pose Regression – 3D-APR) poprawia dokładność metod lokalizacji punktów charakterystycznych twarzy bazujących na Kaskadowej Regresji Kształtu (Cascaded Shape Regression – CSR);
3. Zaproponowana mapa cieplna punktów charakterystycznych pozwala na przekazywanie informacji o ich lokalizacji pomiędzy etapami przetwarzania sieci neuronowej w sposób wizualny, co poprawia dokładność dopasowania twarzy.

Recenzent nie ma zastrzeżeń do sformułowania celu i tezy rozprawy.

Powyższe tezy zostały udowodnione w sposób eksperymentalny. Wysoka odporność Głębokiej Sieci Punktów Twarzy jest udowodniona przez eksperymenty na bazie 300-W Public, uzyskując stopę niepowodzenia niższą o ponad 70% w stosunku do wcześniejszych metod. Poprawa jakości lokalizacji charakterystycznych punktów twarzy przy użyciu metody 3D-APR została udowodniona eksperymentalnie (rozdział 4.6.2) w kontekście metody KRFWS LBF i bazy 300-W Common. Uzyskany wzrost wyniósł 3.5% w stosunku do metody pozbawionej etapu 3D-APR. Zwiększenie skuteczności dopasowania twarzy poprzez zastosowanie mapy cieplnej do przekazywania informacji w sposób obrazowy również zostało udowodnione eksperymentalnie (rozdział 4.7.2). Obliczony współczynnik AUC zwiększył się o 8.3% i 12% odpowiednio w stosunku do przetwarzania bez mapy cieplnej i bez mapy cieplnej i mapy cech. Nie jest niestety określone, dlaczego próg błędu α w metodzie obliczania AUC został zwiększony do 0.08, podczas gdy w konkursie Menpo był on ustalony na 0.05.

Poza szczegółowym opisem proponowanych metod praca zawiera także wyniki dość dużej liczby testów i porównań z innymi metodami na powszechnie uznanych danych benchmarkowych, oraz opis dwóch wybranych zastosowań, które zostały zrealizowane w praktyce.

W recenzowanej rozprawie można wskazać zarówno wątki teoretyczne, jak i eksperymentalne. Do pierwszej grupy można zaliczyć analizę istniejących algorytmów lokalizacji i śledzenia charakterystycznych punktów twarzy. Analiza została uzupełniona wynikami eksperymentów przeprowadzonych przez Autora i twórców oryginalnych metod. Z przedstawionych w części teoretycznej wniosków wynikają kierunki dalszych prac, polegających na opracowaniu nowych algorytmów. Autorskie algorytmy są w dużej mierze rozwinięciem istniejących wcześniej podejść. I tak w przypadku algorytmu FAKRFWS, bazującego na potoku przetwarzania składającym się z APR, 3D-APR i CSR, nowością jest zamiana transformacji afinicznej na transformację podobieństwa na etapie APR oraz zamiana algorytmu regresyjnych drzew decyzyjnych na algorytm K-klastrowych Regresyjnych Drzew Decyzyjnych z Ważonym Dzieleniem (K-cluster Regression Forests With Weighted Splitting). Dodatkowo, na ostatnim etapie zamianie podlegał sposób odbioru cech – z cech gradientowych na piramidę HOG (PHOG). W przypadku sieci DAN, struktura sieci bazuje w swej koncepcji na sieci VGG (Simonyana i Zissermana) wykorzystanej w konkursie ImageNet ILSVRC 2014, uzupełnioną o dodatkowe macierze na wyjściu każdego z etapów przetwarzania – tj. mapę cieplną prezentującą zakres zmienności pozycji punktów charakterystycznych twarzy i macierz cech całego obrazu twarzy (w przeciwieństwie do typowych map przechowujących jedynie informację w okolicach punktów charakterystycznych).

W pracy przedstawiony został oryginalny dorobek Autora zarówno po stronie teoretycznej - dokonanie przeglądu literatury i szczegółowej analizy oraz sformułowanie nowych algorytmów, jak również po stronie doświadczalnej - przebadanie zrealizowanych autorskich algorytmów na uznanym zestawie danych benchmarkowych.

W opinii recenzenta bardzo duże znaczenie ma przedstawiony dorobek eksperymentalny. Potwierdza on wysoką skuteczność opracowanych podejść, a bezkompromisowo dobrany zestaw danych testowych (jako jednych z bardziej wymagających) dowodzi jakości przedstawionych algorytmów.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle) świadczący o dostatecznej wiedzy autora. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Rozprawa zawiera 137 pozycji literaturowych, w tym 7 prac Autora, Promotora i Promotora pomocniczego. Literatura obejmuje zarówno klasyczne pozycje w dziedzinie, jak i najnowsze doniesienia naukowe (cytowane są publikacje z okresu 1901-2018). Pokryty zakres tematyczny jest bardzo szeroki i dobrze umiejscawia recenzowaną rozprawę.

Rozprawa obejmuje omówienie źródeł literatury w kilku grupach zagadnień (metod lokalizacji i śledzenia punktów charakterystycznych, zbiorów benchmarkowych i aplikacji podobnych metod). Obszerny fragment przeglądu literatury zawarty jest w Rozdziale 2 i dotyczy różnych algorytmów lokalizacji i śledzenia punktów charakterystycznych na obrazach twarzy. Autor skupił się na najważniejszych algorytmach z tego zakresu, tj. Active Shape Models, Active Appearance Models, Constrained Local Models, Cascaded Shape Regression oraz Deep Neural Networks. Dokonana w pracy

analiza jest więc cennym źródłem informacji na ten temat.

W pracy nie zabrakło również przeglądu benchmarkowych zestawów testowych, na których przykładzie Autor omówił typowe problemy, takie jak trudna inicjalizacja algorytmu lokalizacji punktów oraz kłopoty ze znaczącymi zmianami w ekspresji i ustawieniu twarzy. Uwzględnił standardowe bazy wykorzystywane w badaniach z dziedziny analizy statycznych obrazów twarzy, tj. Labeled Face Parts in the Wild (LFPW), Helen Facial Feature Dataset (HELEN), Annotated Face in-the-wild (AFW), 300 Faces in-the-Wild Challenge (300-W), WIDER Face, Pointing'04 i Menpo Facial Landmark Localisation in-the-Wild Challenge (Menpo) oraz sekwencji video, tj. 300 Videos in the Wild Challenge (300-VW).

Podsumowując, należy ocenić, że analiza źródeł została sporządzona z dużą starannością i dobrze oddaje stan wiedzy z zakresu omawianych zagadnień na moment złożenia pracy.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Autor, po przedstawieniu w pierwszym rozdziale założeń i celów oraz wprowadzeniu do tematyki rozprawy, w Rozdziale 2 dokonał przeglądu literatury. W Rozdziale 3 przedstawił opis opracowanych, autorskim metod, których testy zaprezentował w Rozdziale 4. Przykładowe zastosowanie opracowanych algorytmów zostały umieszczone w Rozdziale 5. Rozdział 6 prezentuje krótko podsumowanie, dyskusję też oraz opis ewentualnych dalszych prac.

W najważniejszym, z punktu widzenia pracy naukowej, Rozdziale 3, Autor zaprezentował rozważania, które na drodze analizy porównawczej z istniejącymi metodami doprowadziły do opracowania nowych algorytmów lokalizacji i śledzenia punktów charakterystycznych na obrazach twarzy. Droga prowadząca do uzyskania ostatecznego rozwiązania jest, w tym przypadku, ewolucyjna. Proponowane rozwiązania bazują na istniejących algorytmach, które zostały zmodyfikowane i/lub wzbogacone o nowe elementy, co znacząco poprawiło ich skuteczność.

Autor porównywał wyniki dla danych testowych pochodzących z baz 300-W i Menpo, stosując opisane w Rozdziale 4.3 miary jakości i efektywności takie jak średni błąd całkowity (mean absolute error), pole pod krzywą CED (AUC) i odsetek błędnych dopasowań (failure %). Autor uwzględnił również uzyskiwane czasy przetwarzania i wybrane odległości normalizujące, tj. inter-pupil distance i inter-ocular distance. W kontekście wcześniejszych metod, uwzględnił ograniczony zestaw punktów charakterystycznych twarzy (49 zamiast 68), aby porównanie było bardziej miarodajne.

W dysertacji brak jest, niestety, głębszej analizy problemów napotkanych się podczas badań, tzn:

1. Dlaczego FAKRFWS nie radzi sobie ze znaczącymi zmianami ekspresji w rejonie ust (np. rys 4.8)?
2. Czy sekwencja przetwarzania APR + 3D-APR jest wynikiem badań, czy tylko intuicyjnego

przecucia (rys 3.1)? Czy nie można by wyeliminować etapu APR i zamiast niego wprowadzić 3D-APR uwzględniającą pełen zakres przekształceń (również na płaszczyźnie)?

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Do niedawna, większość skutecznych metod lokalizacji punktów charakterystycznych twarzy bazowała na algorytmach z grupy Constrained Local Models (CLM), feature-based Active Shape Models (ASM) i Active Appearance Models (AAM), jak również na architekturach kaskadowych regresji (np. Supervised Descent Method, Explicit Shape Regression itp.). Z drugiej strony, najskuteczniejsze metody śledzenia ww. punktów bazowały na metodach CLM, kaskadowej regresji i modelach z grupy Deformable Part-based Models (DPM). Po pewnym czasie zaczęły pojawiać się w literaturze metody, które łączyły płytkie sieci CNN i rekurencyjne sieci RNN (np. Mnemonic Descent Method). Najnowsze algorytmy łączą gęste metody regresji kształtu i MDM.

W kontekście szybkiego rozwoju metod głębokiego uczenia, najbardziej wartościowym osiągnięciem przedstawionym w rozprawie jest zaproponowany algorytm DAN. Wynika to z nowatorskiego podejścia do rodzaju i liczby przekazywanych informacji na kolejnych etapach uczenia sieci (wizualna reprezentacja punktów charakterystycznych w postaci mapy cieplnej i globalna mapa cech obrazu twarzy). Jego skuteczność została potwierdzona w 2017 roku, kiedy to podczas konkursu 2nd Facial Landmark Localisation Competition – The Menpo BenchMark (w ramach prestiżowej konferencji Computer Vision and Pattern Recognition), zajął trzecie miejsce.

Do kolejnych, wartościowych wyników uzyskanych przez Autora, można zaliczyć algorytm FAKRFWS, który bazuje na bardziej „konwencjonalnym” podejściu do tematyki lokalizacji punktów charakterystycznych oraz połączenie metody APR i metody 3D-ARP w jeden potok przetwarzania, co pozwoliło na zmniejszenie problemów z osiąganiem lokalnych minimów przy poszukiwaniu punktów charakterystycznych twarzy.

Podczas swojej pracy naukowej, Autor opublikował (we współautorstwie) 10 artykułów, z czego 8 pośrednio lub bezpośrednio związanych z tematyką dysertacji. Trzy z nich są zamieszczone w czasopiśmie z tzw. listy A MNiSW. Wyniki osiągnięte przez Autora są konkurencyjne w stosunku do innych, zaprezentowanych na wiodących konferencjach i zamieszczonych w czasopiśmie uwzględnionych w JCR.

Dokonane przez Autora wstępne testy wydajnościowe opracowanych algorytmów FAKRFWS i DAN pokazują, że ich potencjał wdrożeniowy jest bardzo wysoki co jest dodatkowym, wartościowym i oryginalnym osiągnięciem twórczym. W dysertacji nie można niestety znaleźć informacji na temat czasu tworzenia modeli wykorzystujących opracowane metody. Dobór parametrów w praktycznych implementacjach jest również potraktowany bardzo skrótowo. Wydaje się, że dokładniejszy opis tych elementów pozwoliłby na jeszcze trafniejszą ocenę możliwości praktycznego zastosowania opracowanych metod.

Warto zauważyć, że rozwój technologii badanych przez Autora jest obecnie bardzo szybki. Wzrost

zainteresowania metodami głębokiego uczenia jest spowodowany dostępnością odpowiednich algorytmów i wydajnych struktur sprzętowych. Niestety Autor w swej pracy potraktował tę tematykę trochę „po macoszemu”. W swych analizach skupił się na metodach detekcji punktów charakterystycznych, które są podstawą do modelowania twarzy. Jednak równolegle rozwijane są metody, które traktują obraz holistycznie i pozwalają na realizację algorytmów o podobnym zastosowaniu (np. tzw. face-swap), bez kłopotliwego procesu lokalizacji i śledzenia punktów. Chodzi głównie o autoenkodery bazujące na głębokich sieciach neuronowych, które pozwalają na transformowanie jednej twarzy w drugą, przy uwzględnieniu zmian oświetlenia, ekspresji i oczywiście parametrów geometrycznych. Brak jest w rozprawie informacji o algorytmach związanych z tzw. DeepFakes (DeepApp), które wykorzystują frameworki TensorFlow i Keras oraz bardzo duże zbiory obrazów twarzy udostępniane przez wyszukiwarki typu Google Image Search. Ich skuteczność, obserwując zalew materiałów publikowanych w różnych serwisach internetowych, jest bardzo wysoka. Podczas obrony pracy doktorskiej, oczekiwałbym na komentarz Autora w tej kwestii.

5. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

Strona redakcyjna rozprawy stoi na bardzo wysokim poziomie. Użyty język (angielski) jest poprawny i jednoznaczny. Widać w tym aspekcie biegłość i duże doświadczenie Autora w przygotowywaniu publikacji naukowych.

Sposób prowadzenia wywodu w rozprawie jest na dobrym poziomie. Kolejność prezentacji treści jest poprawna. We wprowadzeniu (Rozdział 1) Autor przedstawił motywację do podjęcia prac oraz sformułował cel, zakres i tezę pracy. W Rozdziale 2 przedstawił istniejące metody lokalizacji charakterystycznych punktów twarzy, ze szczególnym uwzględnieniem metod ASM, AAM, CLM, CSR i DNN. W Rozdziale 3 Autor dokonał prezentacji własnych algorytmów: Face Alignment using K-cluster Regression Forests With Weighted Splitting i Deep Alignment Network. Kolejny rozdział (4) zawiera wyniki eksperymentów na benchmarkowych zestawach obrazów (m.in. 300-W i 300-VW). Rozdział 5 prezentuje dwa wybrane zastosowania opracowanych metod (video-okulografia i rzeczywistość rozszerzona). Dysertację kończy Rozdział 6, zawierający podsumowanie.

Autor w swojej pracy prawidłowo prowadzi narrację i poprawnie konstruuje dowody przyjętych na początku tez. Mimo dużej złożoności opisywanych metod potrafi zwrócić uwagę odbiorcy na najistotniejsze fragmenty. W niektórych miejscach rozprawy następuje bardzo drobiazgowo przedstawienie rozmaitych aspektów i szczegółów omawianych metod i eksperymentów. Dowodzi to głębokiemu zrozumieniu problemów podejmowanych w dysertacji.

6. Jakie są słabe strony rozprawy i jej główne wady?

Przedstawiona dysertacja jest dowodem na solidnie wykonaną pracę koncepcyjną i eksperymentalną. Trudno wytknąć jej znaczące wady czy też słabe strony.

Do uwag o stosunkowo niewielkim ciężarze gatunkowym, w obszarze teorii i praktyki, zaliczam:

1. brak szczegółów dotyczących złożoności obliczeniowej opracowanych metod;
2. brak inżynierskiego podejścia do opisy (projektu) opracowanych programów będących implementacją opracowanych algorytmów – chodzi tu o kwestie związane z czasem uczenia struktury drzewiastej i sieci neuronowej, zapotrzebowaniem na pamięć operacyjną, zależnością czasu obliczeń od liczby twarzy uczących, czasu wykonania w zależności od rozdzielczości zdjęcia wejściowego itp. Aspekty te powinny być, w pracy z dziedziny nauk technicznych, szerzej omawiane;
3. brak uzasadnienia niektórych wartości w przyjętych metodach, np. dyskusja parametrów dotyczących modeli regresyjnych (liczba liści w drzewach, liczba drzew) – Rozdział 4.2.1, próg błędu w metodzie AUC (p. 1 niniejszej recenzji) – Rozdział 4.4,
4. brak szerszego omówienia wybranych problemów zaobserwowanych podczas badań, dotyczących genezy powstawania błędów (p. 3 niniejszej recenzji);
5. brak dyskusji różnic w wynikach skuteczności opracowanych metod w zależności od przyjętej odległości normalizującej (inter-pupil distance i inter-ocular distance) – Tabela 4.2;
6. wybrane uwagi dyskusyjne w kontekście innych proponowanych metod i zastosowań, zaprezentowane w p.4 niniejszej recenzji.

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Metody wykrywania i śledzenia punktów charakterystycznych na obrazach twarzy są dość złożonymi algorytmami. Z drugiej strony, istnieje wiele rozwiązań w tej dziedzinie, które rozwijane są od kilkunastu już lat. Jak pokazuje przegląd literatury, część z prostszych problemów nie stanowi już wyzwania naukowego. Z drugiej strony, Autor pokazał w swojej pracy, że pewne problemy dotyczące w szczególności znaczących zmian ekspresji lub orientacji twarzy są cały czas nie do końca rozwiązane. Wobec dynamicznego rozwoju technik widzenia komputerowego, powstawanie prac w tym zakresie ma fundamentalne znaczenie dla współczesnych nauk technicznych, zwłaszcza w kontekście komunikacji człowiek-komputer. Implementacja takich algorytmów w urządzeniach codziennego użytku wymaga wiedzy interdyscyplinarnej, a w przypadku praktycznego zweryfikowania oraz zastosowania tych metod, może dawać unikalne możliwości aplikacyjne.

Praca zawiera także przykłady praktycznego zastosowania opracowanych algorytmów. Pierwszy z nich to prosta aplikacja FaceSwap, która służy do „podmieniania” obrazu twarzy w strumieniu video (Repozytorium GitHub zawiera kody źródłowe, a serwis YouTube przykłady animacji). Drugi to BioWiz - framework do rozpoznawania twarzy opracowany dla Polskiej Policji (Autor nie podaje więcej szczegółów na ten temat). Trzeci to tzw. Character Animation – wspólny projekt ze start-upem ze Stanów Zjednoczonych, ukierunkowany na bez-znacznikowe rejestrowanie ruchu i animację postaci (Autor również nie podaje innych szczegółów na temat tego projektu).

W dalszej części, poświęconej dokładniejszemu opisowi wybranych aplikacji, Autor skupia się na

dwóch innych projektach, które wykorzystują jego osiągnięcia na polu lokalizacji i śledzenia punktów charakterystycznych twarzy. Pierwszy z nich dotyczy video-okulografii, natomiast drugi – rzeczywistości rozszerzonej. W projekcie związanym z okulografią, algorytmy Autora zostały zastosowane do oceny ruchów gałek ocznych zarejestrowanych przy pomocy prostej kamery typu webcam. Rozwiązanie to eliminuje potrzebę zastosowania drogich okulografów i sprawdza się nawet w przypadku znacznych ruchów głowy charakterystycznych dla wielu chorób neurodegeneracyjnych (np. choroby Parkinsona). Drugie zastosowanie, nazwane HoloFace, jest udaną próbą integracji metod Autora w systemie rzeczywistości rozszerzonej. Algorytm detekcji i śledzenia punktów charakterystycznych został w tym przypadku zastosowany do nakładania na obserwowane twarze modeli 3D generowanych w czasie rzeczywistym (przy uwzględnieniu perspektywy oraz zmian w ich skali oraz ekspresji).

Dowodem na znaczącą jakość uzyskanych wyników jest fakt, iż zostały one opublikowane na konferencjach międzynarodowych afiliowanych w IEEE i SPIE oraz w renomowanych czasopismach takich jak IET Computer Vision i IEEE Signal Processing Letters. To dostateczny dowód i potwierdzenie wysokiej wartości wykonanych prac.

8. Do której z następujących kategorii Recenzent zalicza rozprawę:

- a) nie spełniająca wymagań stawianych rozprawom doktorskim przez obowiązujące przepisy
- b) wymagająca wprowadzenia poprawek i ponownego recenzowania
- c) spełniająca wymagania
- d) spełniająca wymagania z wyraźnym nadmiarem

e) wybitnie dobra, zasługująca na wyróżnienie

Podsumowując recenzję, mogę jednoznacznie stwierdzić, że mimo pewnych uwag dyskusyjnych o drugorzędnym znaczeniu, przedstawiona rozprawa Pana mgr inż. Marka Kowalskiego spełnia wymagania ustawowe stawiane rozprawom doktorskim i wnioskuję o dopuszczenie Autora do dalszych etapów przewodu doktorskiego.

Jednocześnie, biorąc pod uwagę uzyskane bardzo dobre wyniki naukowe, konkurencyjne do najlepszych metod state-of-the-art, stawiam wniosek o wyróżnienie rozprawy.



Prof. dr hab. inż. Bogusław Cyganek
Wydział Informatyki, Elektroniki i Telekomunikacji
Katedra Elektroniki
Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

Recenzja rozprawy doktorskiej
mgra inż. Marka Kowalskiego
pt. „*Localization and tracking of facial landmarks in images and*
***video sequences*”**

Wstęp

Przedmiotem recenzji jest rozprawa doktorska Pana mgra inżyniera **Marka Kowalskiego** zatytułowana „*Localization and tracking of facial landmarks in images and video sequences*”, opracowana na Wydziale Elektroniki i Technik Informacyjnych Politechniki Warszawskiej w roku 2018. Promotorem pracy jest prof. dr hab. inż. Władysław Skarbek. Praca została napisana w języku angielskim.

Recenzja zorganizowana jest w postaci odpowiedzi na osiem zasadniczych pytań dotyczących rozprawy doktorskiej.

1. *Jakie zagadnienie naukowe jest rozpatrzone w pracy (tezie rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?*

Praca Pana magistra Marka Kowalskiego dotyczy *opracowania nowych metod lokalizacji i śledzenia punktów charakterystycznych twarzy*. Jako główne osiągnięcie, autor przedstawił obszerny opis dwóch opracowanych przez siebie metod:

- Metoda z k-klastrowymi drzewami regresyjnymi z ważonym podziałem.
- Metoda głębokiej sieci punktów twarzy.

Metody te zostały opracowane, wykonane oraz gruntownie przetestowane przez Autora. Oprócz wyżej wymienionych głównych celów badawczych, w rozprawie zostały podane trzy tezy (Rozdział 1.3), które następnie zostały wykazane w treści pracy i w podsumowaniu (Rozdział 6.1).

Problem naukowy, którego rozwiązania podjął się Doktorant dotyczy komputerowej metody śledzenia punktów charakterystycznych twarzy, w rezultacie czego możliwe jest określenie położenia głowy oraz kierunku patrzenia obserwowanej osoby. Metoda taka stanowi przeważnie kluczowy element pośredni we wszelkiego rodzaju automatycznych systemach wnioskujących o zachowaniu człowieka na podstawie detekcji oraz śledzenia ułożenia jego twarzy.. Systemy tego typu znajdują liczne

zastosowania, na przykład w medycynie, kognitywistyce, psychologii, w systemach wsparcia i kontroli kierowców, w modułach komunikacji człowieka z komputerem, jak również w grach komputerowych. W pracy Autor podaje dwa praktyczne systemy wykorzystujące opracowane przez niego metody śledzenia punktów charakterystycznych twarzy. Pierwszy to system okulograficzny, pozwalający m.in. na wyznaczanie ruchów gałek ocznych w celach diagnostycznych. Drugi, o nazwie HoloFace, to system do rozszerzonej-rzeczywistości (ang. *augmented reality*), umożliwiający m.in. wytwarzanie różnorodnych efektów graficznych z użyciem dynamicznych twarzy. Uwzględniając powyższe, uważam że problem badawczy podjęty przez Doktoranta p. mgra inż. Marka Kowalskiego ma znaczenie zarówno badawcze, jak i aplikacyjne. Charakter pracy – teoretyczno-doświadczalny.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle) świadczący o dostatecznej wiedzy autora? Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

Należy jednoznacznie stwierdzić, że rozprawa doktorska p. mgra inż. Marka Kowalskiego świadczy o jego głębokiej wiedzy w dyscyplinie informatyka, a w szczególności w dziedzinach widzenia komputerowego oraz uczenia maszynowego. Opracowane przez niego metody są na najwyższym światowym poziomie, a uzyskane wyniki eksperymentalne plasują je wśród ścisłej czołówki rozwiązań światowych, tzw. *state-of-the-art*.

Osiągnięcia te nie byłyby możliwe bez dogłębnej analizy najnowszych źródeł i opracowań światowych. Zaprezentowany przegląd źródeł oraz podobnych rozwiązań światowych świadczy o doskonałej znajomości tematu przez Doktoranta. Wnioski z przeglądu tych rozwiązań przyczyniły się do opracowania dwóch wysoce wydajnych metod śledzenia punktów charakterystycznych twarzy.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Rozprawa doktorska Pana mgra inż. Marka Kowalskiego dotyczy opracowania wysoce skutecznych metod śledzenia punktów charakterystycznych twarzy. Skuteczność oznacza tutaj zarówno wysoką dokładność odpowiedzi, jak również jednoczesne działanie umożliwiające pracę w tzw. trybie czasu rzeczywistego. Obydwa powyższe aspekty decydują o przydatności metody. Należy jednoznacznie stwierdzić, że **wyżej przedstawione cele zostały osiągnięte** w postaci opracowania, jak i realizacji dwóch alternatywnych metod śledzenia punktów charakterystycznych twarzy. Obydwie te metody zostały następnie gruntownie przetestowane. Wyniki przeprowadzonych testów, świadczą o zrealizowaniu głównego celu projektu. Otrzymana dokładność odpowiedzi, jak również czasy działania są porównywalne z najlepszymi rozwiązaniami światowymi. Należy też wspomnieć, że obydwie metody zostały opublikowane w renomowanych periodykach naukowych o zasięgu światowym.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Do najważniejszych osiągnięć naukowych autora, opisanych w recenzowanej pracy i stanowiących oryginalny dorobek zaliczyć można zaproponowanie, opracowanie teoretyczne, wykonanie oraz przetestowanie dwóch różnych metod wyznaczania punktów charakterystycznych twarzy. Poniżej przedstawiona została krótka charakterystyka obydwu rozwiązań.

1. Metoda lokalizacji punktów charakterystycznych twarzy z wykorzystaniem k -klastrowego lasu drzew regresyjnych z ważonym podziałem, (nazwanej w pracy ang. *Face Alignment using K-cluster Regression Forests With Weighted Splitting, FAKRFWS*) – w metodzie tej zaproponowano nowe podejście polegające na wykorzystaniu drzew regresyjnych w postaci k -klastrowych lasów regresyjnych z ważonym podziałem, afinicznej metody regresji 3D (ang. *3D Affine Pose Regression*) zmniejszającej wpływ początkowego arbitralnego ułożenia głowy, jak również deskryptorów obrazu PHOG (ang. *Pyramid of Histograms of Oriented Gradients*). W metodzie tej udoskonalono jej wcześniejszą wersję, działającą na „zwykłych” lasach regresyjnych i wcześniej używanych deskryptorów lokalnych wzorców binarnych LBP (ang. *Local Binary Patterns*). Znacznie została udoskonalona faza inicjalizacji metody, która musi brać pod uwagę początkowe arbitralne ułożenie głowy człowieka. W tym celu Doktorant najpierw zaproponował użycie transformacji 2D do transformacji początkowego kształtu twarzy w celu kompensacji niedokładności początkowego oszacowania rozmiaru oraz obrotu. Przekształcenie to zostało następnie rozszerzone do transformacji typu 3D, która dodatkowo umożliwia kompensację dowolnego obrotu twarzy w przestrzeni 3D, a nie tylko obrotu w obrębie jednej płaszczyzny. Jednak główna modyfikacja wyżej opisanego schematu, zaproponowana przez p. Marka Kowalskiego, to zastosowanie k -klastrowego lasu regresyjnego połączonego z mechanizmem podziału drzew na podstawie wag, jak również zastąpienie LBP bardziej wydajnymi w tym przypadku cechami PHOG. W metodzie tej, przestrzeń danych wejściowych podzielona jest na k klastrów, które lokalnie minimalizują funkcję strat przypisaną do drzewa (w tym przypadku jest to suma kwadratów różnic składowych). Każdy z k podziałów dokonywany jest za pomocą k -klasowego klasyfikatora z wektorami wspierającymi SVM (ang. *Support Vector Machine*). Dzięki temu możliwy jest podział wykorzystujący nie tylko jednowymiarowe, a wielo-wymiarowe dane, takie jako HOG czy PHOG. Dodatkowo, z każdą daną związana jest pewna waga, wykorzystywana przy każdej klasyfikacji związanej z nowym podziałem przy konstrukcji drzewa. Waga taka pozwala uwypuklić role danych, których ewentualna nieprawidłowa klasyfikacja (tj. niedoskonałe przypisanie danych do klastrów), prowadzi do znacznego zwiększenia błędu klasyfikacji.
2. Opracowanie metody z wykorzystaniem głębokiej sieci neuronowej (ang. *Deep Alignment Network, DAN*) – w metodzie tej kluczową rolę odgrywają odpowiednio uczone warstwy neuronowe, tworzące tzw. głęboką sieć neuronową. Sieci te, znane od końca lat 90-tych, dzięki nowym osiągnięciom teoretycznym, jak również znacznie większym mocom obliczeniowym współczesnych komputerów, zrewolucjonizowały dziedzinę uczenia maszynowego, umożliwiając klasyfikację z dotychczas nieosiągalną dokładnością. Sieć opracowana przez p. mgra inż. Marka Kowalskiego składa się z wielu warstw, z których każda kolejna doprecyzowuje wartości położenia charakterystycznych punktów twarzy, obliczone i przekazane przez warstwę poprzedzającą. Jednakże w przeciwieństwie do innych metod tego typu, na

wejście sieci podawane są całe obrazy twarzy, a nie np. zbiory lokalnych pod-obrazów. Taki tryb działania jest korzystny, gdyż umożliwia dopasowanie się sieci do większej ilości wariantów ułożenia twarzy w warunkach rzeczywistych. To z kolei ułatwia sam proces inicjalizacji sieci. Jednakże podawanie całych obrazów twarzy stało się możliwe dzięki zaproponowanemu mechanizmowi generowania tzw. map intensywnego występowania cech (ang. *heatmaps*). Mapy te dostarczają dodatkową informację do każdej kolejnej warstwy sieci (rodzaj lokalnej wagi) na temat występowania punktów charakterystycznych twarzy, wyznaczonych w poprzedzającej warstwie sieci. Jest to więc rodzaj lokalnego mechanizmu wizualnej uwagi (ang. *visual attention*), zapobiegający również wpadaniu metody w tzw. lokalne minima. Mechanizm ten ma też wiele wspólnego z tzw. przestrzeniami skal. W ten sposób, każda z kolejnych warstw sieci otrzymuje aż trzy rodzaje informacji: (1) obraz wejściowy twarzy, zakrzywiony geometrycznie tak aby właściwie dopasować bieżące punkty charakterystyczne do kształtu kanonicznego, (2) mapę intensywności występowania cech, oraz (3) obraz cech wygenerowany na podstawie gęstych cech poprzedniej warstwy. Sieć operuje w trybie „przesuwu wprzód” (ang. *feed-forward*). Na specjalną uwagę zasługuje zaproponowanie funkcji błędu uwzględniającej geometrię twarzy, wykorzystywanej podczas uczenia sieci, jak również przebadanie aż trzech wariantów trenowania sieci. W rezultacie otrzymane zostały wyniki, które na trudnym zbiorze testowym 300-W plasują metodę wśród najlepszych metod tego typu na świecie.

Opisane metody lokalizacji punktów charakterystycznych twarzy należą do grupy tzw. metod kaskadowej regresji kształtu (ang. *Cascaded Shape Regression*), oryginalnie zaproponowanej w pracy Cao et al. „*Face alignment by explicit shape regression*” i opublikowanej w 2014 roku. Główna idea działania tych metod polega na sukcesywnym uaktualnianiu wstępnego oszacowania modelu w postaci zbioru pozycji punktów charakterystycznych, zaczynając od pewnej wartości startowej aż do osiągnięcia zbieżności po kilku iteracjach. W każdej z iteracji, wyznaczane są specyficzne cechy o obszarach występowania punktów charakterystycznych twarzy. Cechy te są następnie używane w mechanizmie regresyjnym do estymacji kolejnych wartości. Schemat działania wpisuje się więc w kaskadę przekształceń, skąd pochodzi nazwa tej grupy metod.

Powyższe metody są w pewnym sensie komplementarne. Pierwsza z nich, tj. metoda KRFWS, osiąga nieco niższą dokładność, za to działając nieco szybciej. Dzięki temu możliwe jest jej zastosowanie w aplikacjach mobilnych, w których preferowane są rozwiązania nie wymagające zbyt dużych mocy obliczeniowych. Z kolei metoda DAN wpisuje się w grupę metod opartych na głębokich sieciach neuronowych, które stanowią najefektywniejsze rozwiązania uczenia maszynowego ostatnich lat.

Autor przeprowadził dogłębną analizę wyników działania poszczególnych metod i wykazał, że obydwie plasują się w czołówce znanych rozwiązań światowych w tej dziedzinie. Na szczególne wyróżnienie zasługuje tutaj zajęcie przez Doktoranta wraz ze współautorami i metodą DAN trzeciego miejsca w zawodach 2nd *Facial Landmark Localisation Competition – The Menpo BenchMark*, zorganizowanych w roku 2017 podczas konferencji *Conference on Computer Vision and Pattern Recognition* (CVPR 2017), należącej do jednego z najbardziej prestiżowych spotkań naukowców z dziedzin CV oraz PR na świecie.

Wszystkie wyżej wymienione osiągnięcia świadczą o dużej dojrzałości naukowej p. mgra inż. Marka Kowalskiego, który swobodnie porusza się zarówno w dziedzinie widzenia komputerowego, uczenia maszynowego, metod algebry obliczeniowej, jak również w dziedzinie programowania systemów komputerowych.

3. Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

Rozprawa doktorska p. mgra inż. Marka Kowalskiego liczy ponad 120 stron. Napisana jest w języku angielskim. Praca podzielona została na 6 rozdziałów głównych. Bibliografia liczy 137 pozycji literaturowych. Oddzielnie wymienione są prace których głównym autorem lub współautorem jest Doktorant.

Układ pracy uważam za prawidłowy. Praca napisana jest jasno bez błędów językowych, czy też edycyjnych. Uwzględniwszy jednak fakt, że główny opis dwóch prezentowanych przez Doktoranta metod został już opublikowanych w artykułach naukowych, których łączna ilość stron nie przekracza 20, to rozprawa doktorska mogłaby być bardziej zwięzła, podając jedynie krótki opis osiągnięć i pewne tło naukowe do uprawianej dziedziny, jednocześnie bezpośrednio powołując się na merytoryczny opis metod zawarty we wspomnianych pracach naukowych Autora.

Należy dodać, że Doktorant, Pan mgr inż. Marek Kowalski, jest współautorem 10 publikacji naukowych, z których w 5 jest pierwszym autorem. Trzy z wyżej wymienionych zostały opublikowane w znanych czasopismach naukowych posiadających tzw. *Impact Factor* (IF). Z tych, za najważniejsze uważam:

- publikację zatytułowaną *“Face alignment using k-cluster regression forests with weighted splitting”*, autorstwa M. Kowalski oraz J. Naruniec, a opublikowaną w *IEEE Signal Processing Letters* (bieżący IF= 2,528), w roku 2016,
- jak również *„Deep Alignment Network: A convolutional neural network for robust face alignment”*, autorstwa M. Kowalski, J. Naruniec, i T. Trzcinski, opublikowaną na *IEEE Conference on Computer Vision and Pattern Recognition Workshops* w roku 2017.

Wszystkie te publikacje dotyczą dziedziny widzenia komputerowego i uczenia maszynowego wraz z aplikacjami.

4. Jakie są słabe strony rozprawy i jej główne wady?

W pracy nie zauważyłem żadnych błędów merytorycznych, czy też poważniejszych usterek edycyjnych. Praca napisana jest w sposób jasny i zorganizowany, a wszystkie zagadnienia dotyczące istotnych osiągnięć Doktoranta zostały dogłębnie przeanalizowane i opisane. Uwagi, które przedstawiam poniżej mają charakter dyskusyjny, bądź też dotyczą zagadnień wymagających dodatkowych informacji ze strony Doktoranta.

1. Sprzeczne informacje odnośnie użycia LBP (str. 48) – jako trzeci punkt na liście głównych osiągnięć Autor pisze *„a face alignment method based on KRFWS and the Local Binary Feature framework”*, podczas gdy parę zdań wcześniej zaznacza, że zamiast LBP używa PHOG.
2. Nie do końca jest jasne jak dokonywane są podziały drzew w KRF na podstawie klasyfikacji danych wielowymiarowych. Czy za dane wielowymiarowe Autor uważa wektory, czy też ogólniejsze wielowymiarowe obiekty tensorowe?
3. Czy są jakieś różnice w używaniu pakietu LIBLINEAR do implementacji ważonych klasyfikatorów typu SVM w stosunku np. do LibSVM?

4. W metodzie KRFWS, nie jest jasny proces generowania początkowych klastrów.
5. W początkowym opisie KRFWS jest mowa o ogólności drzew i możliwości budowania węzłów o liczbie potomków (ang. *child*) większej niż 2. Jednakże w testach wykorzystywane były wyłącznie 2 węzły, bez sprawdzania czy inna ich ilość byłaby lepsza. Czy Autor ma jakieś przesłanki teoretyczno-eksperymentalne tłumaczące ten wybór?
6. Jaki rodzaj danych typu ground-truth musi być użyty do metody treningowej opisanej w rozdziale 3.1.5? W jaki sposób otrzymywane są te dane typu ground-truth dla tej metody?
7. W problemie śledzenia często wykorzystuje się informację z poprzednich ramek do określenia położenia obiektów w ramce bieżącej (np. mean-shift, filtr Kalmana, itd.). Jak wygląda ten rodzaj przekazywania informacji w przypadku metod śledzenia opisanych w rozprawie?
8. W opisie wyników, np. Tabela 4.1 i dalsze, dobrze byłoby podać rozdzielczość ramek użytych do eksperymentów.
9. Skoro testy są wykonywane wiele razy to nasuwa się pytanie czy i jak powtarzalne są wyniki? Jeśli jest element losowości, to czy można by rozważyć użycie testów statystycznych do porównywania metod. A jeśli tak, to jakich?
10. Wiele detali odnośnie inicjalizacji pozostaje niewiadomych (np. losowa inicjalizacja połączeń sieci, itd.).
11. W rozdziale (4.5) opisana jest metoda detekcji twarzy z wykorzystaniem funkcji pakietu OpenCV. Jaki dokładnie algorytm został użyty? Od jakich parametrów on zależy?
12. W analizie wyników przedstawionych w tabeli 4.5 można by rozważyć użycie statystycznego testu rankingowego do analizy wyników. Podobnie w przypadku analizy wyników przedstawionej w rozdziale (4.6.1), w którym Autor pisze „*KRFWS again obtains the best results, but the difference is less significant*”. Być może jest to mniejsza istotność, ale nasuwa się pytanie w jakim sensie?
13. Obserwując działanie obydwu opracowanych metod, nasuwa się sugestia/pytanie odnośnie ewentualnego ich połączenia. Dobre wyniki uzyskuje się poprzez połączenie metod w tzw. *ensemble*, szczególnie metod uzupełniających się, posiadających duży współczynnik tzw. różnorodności (ang. *diversity*). Czy Doktorant mógłby rozwinąć ten temat? Czy jest to możliwe w przypadku dwóch zaproponowanych metod? Jeśli tak, to czego możemy się spodziewać?
14. Czy jest jakieś podobieństwo między funkcją kosztu użytą w DAN, a tą w metodzie KRFWS?
15. W przypadku tabel 4.7 oraz 4.8 – jaka jest różnica pomiędzy pojęciami „*mean average error*” vs. „*mean error*”?
16. Miary podane w tabeli 4.10 być może wymagają dodatkowego wyjaśnienia, gdyż każda z nich przenosi nieco inną informację; Różna jest więc interpretacja 4.47 vs. 4.30 od 3.17 vs 2.67, itd.
17. Dodatkowego wyjaśnienia wymagają wyniki w tabeli 4.2, szczególnie w kontekście różnych miar „*inter-pupil*” vs. „*inter-ocular*” na zbiorze 300-W, który został oznaczony tylko dla jednej z nich. Z kolei, dla tego samego zbioru testowego, w tabeli 4.3 oraz 4.4 znajdujemy wyniki wyłącznie dla „*inter-ocular*”.
18. Na szczególną analizę i dyskusję zasługuje problem uczenia sieci w trybie całościowym, tzn. *end-to-end*. W wielu przypadkach i przy różnych architekturach sieci tryb ten jest wielce pożądany. Jednakże w tym trybie uczenia Doktorantowi nie udało się uzyskać wystarczająco zadowalających wyników. Co więcej, kolejne istotne prace naukowe raportują duże sukcesy właśnie z wykorzystaniem tego trybu uczenia (np. niedawna praca Lai *et al.*: *Deep Recurrent Regression for Facial Landmark Detection*. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 28, No. 5, May 2018). Bardzo ciekawe byłoby porównanie przez Doktoranta jego wersji sieci DAN w kontekście np. wymienionej pracy, w której również zaproponowana została głęboka sieć regresyjna, której raportowana dokładność na zbiorze 300-W wynosi 4,07.

Tak jak zostało to już zaznaczone we wstępie, powyższe uwagi mają charakter bardziej polemiczny i dotyczą dalszego wyjaśnienia, bądź też rozszerzenia pewnych zagadnień natury naukowej poruszonych w rozprawie.

5. Jaka jest przydatność rozprawy dla nauk technicznych?

Metody opracowane oraz opisane w rozprawie doktorskiej p. mgra inż. Marka Kowalskiego mają istotne znaczenie dla nauk technicznych, zarówno w aspekcie osiągnięć teoretycznych, jak i możliwości aplikacyjnych. Przydatność ta jest pochodną głównych i samodzielnych osiągnięć autora, o których była mowa w poprzednich punktach recenzji. W szczególności, podkreślić należy główne osiągnięcia, czyli opracowanie dwóch metod wyznaczania oraz śledzenia punktów charakterystycznych twarzy, charakteryzujące się dużą dokładnością odpowiedzi porównywalną ze rozwiązaniami światowymi *state-of-the-art*, jak również operującymi w trybie tzw. czasu rzeczywistego.

Opracowane metody mogą znaleźć liczne zastosowania we wszelkiego typu systemach komputerowych wymagających śledzenia ułożenia głowy obserwatora, kierunku patrzenia, jak również ogólniej pojętej komputerowej analizy zachowania człowieka w różnych sytuacjach. Tak jak wykazał Autor, system ten w szczególności może być zastosowany w aplikacjach medycznych oraz w systemach rozrywkowych, filmach oraz grach komputerowych, jak również w aplikacjach mobilnych.

6. Do której z następujących kategorii Recenzent zalicza rozprawę:

- a) nie spełniająca wymagań stawianych rozprawom doktorskim przez zobowiązujące przepisy
- b) wymagająca wprowadzenia poprawek i ponownego recenzowania
- c) spełniająca wymagania
- d) spełniająca wymagania z wyraźnym nadmiarem
- e) wybitnie dobra, zasługująca na wyróżnienie

Podsumowując stwierdzam, że postawione w rozprawie zadania badawcze zostały rozwiązane. Uzyskane rezultaty stanowią oryginalny własny wkład autora rozprawy Pana mgra inżyniera Marka Kowalskiego w rozwój dyscypliny naukowej Informatyka, a w szczególności dziedzin widzenia komputerowego oraz uczenia maszynowego. Wykazał się on przy tym dużą znajomością najnowszych metod i algorytmów przetwarzania obrazów, klasyfikacji danych, doskonałym opanowaniem warsztatu badawczego, a w rezultacie również dojrzałością naukową.

Recenzowaną pracę oceniam jako **spełniającą z nadmiarem** wymagania stawiane rozprawom doktorskim i **zasługującą na wyróżnienie**. Wnioskuje o jej przyjęcie oraz o dopuszczenie Pana magistra inżyniera Marka Kowalskiego do publicznej obrony.

Bożenka Cyg