

Warszawa, dnia 06.06.2018

Autoreferat

- 1. Imię i nazwisko:** Jacek Krzysztof Naruniec
- 2. Posiadane dyplomy, stopnie naukowe – z podaniem nazwy, miejsca i roku ich uzyskania oraz tytułu rozprawy doktorskiej.**

Tytuł zawodowy inżyniera

*Rok 2006,
Wydział Elektroniki i Technik Informatycznych,
Politechnika Warszawska*

Tytuł zawodowy magistra

*Rok 2006,
Wydział Elektroniki i Technik Informatycznych,
Politechnika Warszawska*

*Stopień doktora nauk technicznych,
praca pt. „Metody częstotliwościowo-
przestrzenne w detekcji i śledzeniu
charakterystycznych punktów twarzy”*

*Rok 2010,
Wydział Elektroniki i Technik Informatycznych,
Politechnika Warszawska*

- 3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych/ artystycznych.**

<i>Okres zatrudnienia</i>	<i>Forma zatrudnienia/stanowisko</i>	<i>Miejsce zatrudnienia</i>
2010-2011	asystent naukowy	Politechnika Warszawska Wydział Elektroniki i Technik Informatycznych Instytut Radioelektroniki
2011-2012	umowa zlecenie	Centrum Badań Kosmicznych Polskiej Akademii Nauk
2012-2015	umowa zlecenie	Instytut Maszyn Matematycznych
2011-obecnie	adiunkt	Politechnika Warszawska Wydział Elektroniki i Technik Informatycznych Instytut Radioelektroniki i Technik Multimedialnych

4. Wskazanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.).

a) Tytuł osiągnięcia naukowego:

Jako osiągnięcie naukowe zgłaszam monotematyczny cykl 8 publikacji, obejmujący 7 artykułów i patent, pt.:

„Metody, algorytmy i zastosowania analizy twarzy”

b) autor/autorzy, tytuł/tytuły publikacji, rok wydania, nazwa wydawnictwa, recenzenci wydawniczy,

[H1] Naruniec, J.: „Discrete Area Filters in Accurate Detection of Faces and Facial Features”, Image and Vision Computing, Volume 32, Issue 12, s. 979-993, ISSN 0262-8856, 2014, **35 p., IF. 1,587**

[H2] Kowalski, M., Naruniec, J.: „Face Alignment Using K-Cluster Regression Forests With Weighted Splitting”, IEEE Signal Processing Letters, vol. 23, issue 11, s. 1567-1571, 2016, **30 p., IF. 2,528**

[H3] Kowalski, M., Naruniec, J., Trzeciński T.: „Deep Alignment Network: A convolutional neural network for robust face alignment”. 2017 IEEE International Conference on Computer Vision and Pattern Recognition Workshop, 2017, **15 p.**

[H4] Będkowski, J., Naruniec J.: „On-line Range Registration with GPGPU”, in Opto-Electronics Review, Springer, vol. 21, nr 1, s. 52-62, 2013, **20 p., IF. 1,279**

[H5] Naruniec J., Kowalski M., Daniluk M.: „3D Face Data Acquisition and Modelling Based on an RGBD Camera Matrix”. Conference proceedings: 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Warszawa, Polska, 2015, **15 p.**

[H6] Strupczewski, A., Naruniec J., Mucha K., Czyprynski, B.: „Eye Gaze Tracking Method and Apparatus and Computer-Readable Recording Medium”, US Patent 20,150,293,588 (Samsung Electronics Co.), 2016

[H7] Naruniec, J., Wieczorek, M., Szlufik, S., Koziorowski, D., Tomaszewski M., Kowalski M., Przybyszewski A.: „Webcam-based system for video-oculography”, IET Computer Vision, vol. 11, issue 2, pp 173-180, 2017, **20 p., IF. 0,878**

[H8] Wrobel, K., Doroz, R., Porwik, P., Naruniec, J., Kowalski M.: „Using a Probabilistic Neural Network for lip-based biometric verification”. Engineering Applications of Artificial Intelligence, vol. 64, s. 112-127, 2017, **35 p. IF. 2,894**

Oświadczenia mojego indywidualnego wkładu w powstanie wymienionych wyżej publikacji znajdują się w załączniku 3. Oświadczenia współautorów określające indywidualny wkład każdego z nich zestawiono w załączniku 4.

- c) *omówienie celu naukowego/artystycznego ww. pracy/prac i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania.*

Analizę twarzy można określić działaniami prowadzące do uzyskania określonych informacji na podstawie jej obrazu, takich jak tożsamość osoby, emocje, pozycja punktów szczególnych, poza głowy, struktura trójwymiarowa, zmienność w czasie, czy też punkt skupienia wzroku. Jest to dziedzina bardzo popularna ze względu na fakt, że wiele praktycznych zastosowań nie wymaga interakcji ze strony użytkownika. Co więcej, niektóre algorytmy, takie jak rozpoznawanie twarzy czy emocji, mogą działać bez wiedzy osoby widocznej w obrazie.

Istnieje wiele wyzwań dotyczących analizy twarzy, których rozwiązania nie zapewniają jeszcze skuteczności umożliwiającej komercyjne zastosowanie. Do takich problemów wlicza się np. wyznaczenie punktu fiksacji wzroku w analizie obrazu kamer światła widzialnego. Skuteczna metoda realizująca to zadanie jest wysoce pożądana, np. ze względu na możliwość nieinwazyjnego określenia czasu obserwacji produktu przez klienta. W niniejszym cyklu publikacji ten temat został także przeanalizowany.

W przedstawionym cyklu publikacji przebadano i zaproponowano szereg metod i algorytmów służących do lokalizacji punktów szczególnych twarzy [H1][H2][H3]. Jest to algorytm kluczowy z punktu widzenia precyzji działania pozostałych proponowanych rozwiązań. Stworzone metody lokalizacji punktów szczególnych zostały wykorzystane do realizacji algorytmów analizy twarzy, obejmujących wyznaczenie i śledzenia punktu fiksacji wzroku (patent [H6] opublikowany w ramach współpracy z Samsung R&D), okulografię [H7] i rozpoznawanie tożsamości na podstawie obrazu ust [H8]. Punkty szczególne zostały także wykorzystane w algorytmie rejestracji i modelowania trójwymiarowego twarzy [H4][H5].

Wiele najnowszych algorytmów analizy twarzy charakteryzuje się wysoką złożonością obliczeniową, wynikającą np. z faktu wykorzystania złożonych splotowych sieci neuronowych lub analizy dużych ilości danych zlokalizowanych w przestrzeni trójwymiarowej. Z tego względu, w ramach proponowanych algorytmów, część obliczeń wykonywana jest z wykorzystaniem kart graficznych z procesorami ogólnego przeznaczenia. Temat wykorzystania takich procesorów w przetwarzaniu danych trójwymiarowych został poruszony w publikacji [H4].

Celem prac, których wynikiem jest przedstawiony cykl publikacyjny jest zaproponowanie szeregu oryginalnych metod, oraz skutecznych i szybkich algorytmów analizy twarzy, realizujących zarówno klasyczne zadania przetwarzania obrazów, jak i zadania uczenia maszynowego i rekonstrukcji trójwymiarowej.

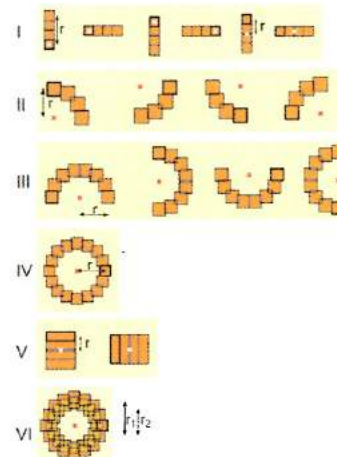
5. **Opis osiągnięcia** wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.).

5.A Algorytmy wykrywania i lokalizacji szczególnych punktów twarzy.

W publikacji [H1] zaproponowany został **nowy algorytm wykrywania punktów szczególnych twarzy**, który **charakteryzuje się** wysoką skutecznością, **przewyższającą** wiele **czołowych metod dostępnych w momencie publikacji**. W pracy opisany został algorytm będący rozszerzeniem metod powstałych w ramach mojej pracy doktorskiej. Cechą istotnie wyróżniającą proponowany algorytm od innych, dostępnych w tym czasie algorytmów, jest **jednoczesna detekcja twarzy wraz z lokalizacją zestawu 15 szczególnych punktów twarzy**. Cel ten uzyskano stosując podejście „od szczegółu do ogółu”, podczas gdy w literaturze dominowało podejście odwrotne – „od ogółu do szczegółu”. W podejściu „od ogółu do szczegółu” w pierwszym kroku wykrywane są wszystkie twarze, a następnie, w ramach znalezionych twarzy, lokalizowane są punkty szczególne.

W proponowanym algorytmie („od szczegółu do ogółu”) w pierwszej kolejności piksele klasyfikowane są do jednej lub wielu z 15 kategorii punktów szczególnych, lub do kategorii „tło”. W drugiej kolejności sklasyfikowane punkty grupowane są w osobne części twarzy - nos, usta, oczy, a na końcu części twarzy grupowane są w osobne twarze. Inną istotną cechą proponowanego algorytmu są **nowe cechy obrazu nazwane *discrete area filters***, co można tłumaczyć jako dyskretne filtry obszarowe. Dyskretny filtr obszarowy złożony jest z zestawu prostokątów tworzących jeden z 6 zdefiniowanych kształtów (rys. 1). Ekstrakcja cech rozpoczyna się od wyznaczenia sumy składowych luminancji pikseli zawierających się w każdym ze zdefiniowanych prostokątów. Otrzymane wartości są następnie normalizowane do przedziału $\langle 0..1023 \rangle$. Wektor powstały z konkatencji wybranego zestawu dyskretnych filtrów obszarowych, jest następnie przekształcany liniowo. Macierz przekształcenia określana jest osobno dla każdego typu punktów charakterystycznych i typu filtrów obszarowych poprzez zastosowanie metod liniowej analizy dyskryminacyjnej. Szczegóły trenowania jak i konstrukcji filtrów można znaleźć w oryginalnej publikacji [H1]. Klasyfikacja cechy realizowana jest poprzez progowanie odległości euklidesowej wyznaczonej cechy do średniej cech wyznaczonych dla przykładów uczących danego typu punktu szczególnego.

Łączenie punktów szczególnych w części twarzy i twarze zrealizowane zostało poprzez wykorzystanie grafów referencyjnych. Dzięki podejściu „od szczegółu do ogółu” algorytm



Rysunek 1: Kształt zaproponowanych sześciu typów dyskretnych filtrów obszarowych wraz z możliwymi obrotami. Czerwony punkt oznacza punkt zaczepienia filtra w obrazie. Źródło: H2.



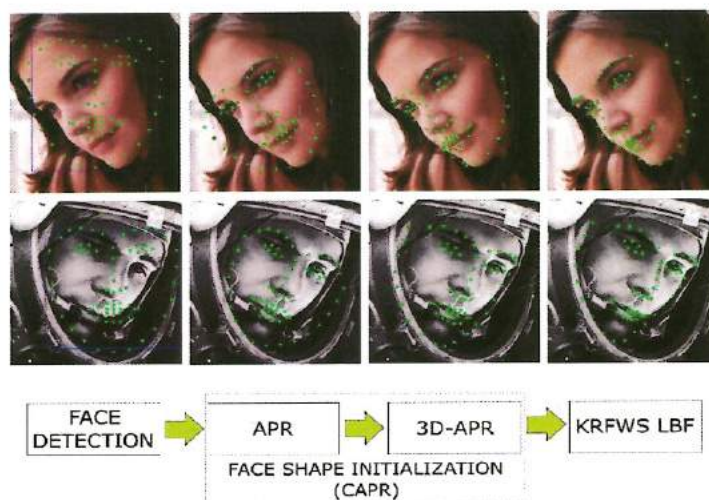
Rysunek 2: Przykłady poprawnych detekcji twarzy za pomocą algorytmu dyskretnych filtrów obszarowych. Źródło: [H1].

może także służyć do detekcji twarzy częściowo przysłoniętych. Pozycje niewidocznych części twarzy są aproksymowane na podstawie grafu referencyjnego i pozycji pozostałych, wykrytych elementów. Rys 2. przedstawia przykłady poprawnych detekcji za pomocą przedstawionego algorytmu.

Biblioteka programistyczna, zawierająca implementację przedstawionego algorytmu, opublikowana została na stronie internetowej, natomiast film demonstrujący detekcję za pomocą tego algorytmu został opublikowany w serwisie YouTube, gdzie został wyświetlony ponad 19000 razy.

W 2013 roku Cao i in. opublikowali pracę pod tytułem „*Face alignment by explicit shape regression*”, której idea stanowiła podstawę wielu nowych metod lokalizacji szczególnych punktów twarzy (ang. *face alignment*), o niespotykanej dotąd skuteczności. Algorytm poprzedzony jest detekcją twarzy. Następnie, w obszarze obejmującym wykrytą twarz, ustalana jest początkowa pozycja każdego z punktów szczególnych. Kształt początkowy ustalany jest jako średni kształt wszystkich manualnie oznaczonych twarzy wykorzystanych w procesie uczenia. W kolejnych iteracjach algorytmu, pozycje punktów są uszczegóławiane tak, aby dopasować je do twarzy widocznej na zdjęciu. Każda iteracja realizowana jest poprzez funkcję regresji, której parametrem jest aktualny deskryptor obrazu, natomiast jej wynikiem są przesunięcia poszczególnych punktów charakterystycznych względem ich aktualnych pozycji. Kolejne funkcje regresji tworzą tzw. kaskadę regresorów.

W pracy [H2] idea ta została rozszerzona o szereg elementów. **Nasze badania wykazały, że regresja z wykorzystaniem drzew losowych i maszyny wektorów nośnych (*Support Vector***



Rysunek 3: Schemat przedstawionego w [H2] algorytmu lokalizacji punktów szczególnych twarzy.

Machine) prowadzi do istotnego zwiększenia precyzji, zastępując dwustopniową regresję łączoną (ang. *two-step boosted regression*). Cechy kontrastowe zastąpione zostały cechami *Pyramid Histogram of Oriented Gradients (PHOG)*, które określają histogram rozkładu kierunku krawędzi w wielu skalach obrazu. **Innowacyjnym elementem pracy jest także wprowadzenie wagowania przykładów w procesie uczenia klasyfikatorów**, które wykorzystywane były jako narzędzie podziału w kolejnych węzłach drzew losowych. **Istotną poprawę działania oryginalnego algorytmu uzyskano także dzięki zastosowaniu dwustopniowej metody inicjalizacji początkowej pozycji punktów szczególnych.**

Zaproponowana metoda lokalizacji szczególnych punktów w obrazie realizowana jest w następujących krokach:

- Wykryj twarz w obrazie (dowolny algorytm detekcji twarzy) i określ początkową pozycję punktów szczególnych twarzy poprzez dopasowanie średniego zestawu tych punktów do wyznaczonego obszaru twarzy.
- Zmodyfikuj początkową pozycję punktów szczególnych poprzez afiniczne przekształcenie wzorcowego zestawu punktów szczególnych (algorytm *Affine Pose Regression*) umożliwiającego liniowe dopasowanie punktów do twarzy w obrazie. Parametry przekształcenia określone są na podstawie regresji deskryptora obrazu.
- Dokonaj trójwymiarowego (translacja i rotacja) przekształcenie kształtu w celu dopasowania aktualnej pozycji punktów do twarzy w obrazie (algorytm *3D Affine Pose Regression*). Parametry przekształcenia określone są na podstawie regresji deskryptora obrazu.
- Wyznacz dokładną pozycję punktów szczególnych stosując proponowaną, iteracyjną metodę regresji opartą na drzewach losowych i klasyfikatorze SVM (algorytm *K-cluster Regression Forests with Weighted Splitting - KRFWS*).

Wejściem każdego z proponowanych algorytmów (*Affine Pose Regression*, *3D Affine Pose Regression*, *KRFWS*) jest deskryptor utworzony z konkatencji cech *PHOG* wyznaczonych z obrazu w aktualnych pozycjach punktów szczególnych twarzy. Kroki te ilustruje rysunek 3.

Dzięki zastosowanym technikom **otrzymano wyniki przewyższające precyzją lokalizacji istniejące, najlepsze algorytmy**. Wyniki lokalizacji weryfikowane były na jednej z najpopularniejszych baz służących do badania algorytmów lokalizacji punktów szczególnych twarzy – 300W. Jest to baza zawierająca twarze sfotografowane w warunkach naturalnych, często przy różnorodnym oświetleniu, pozie głowy i ekspresji twarzy.

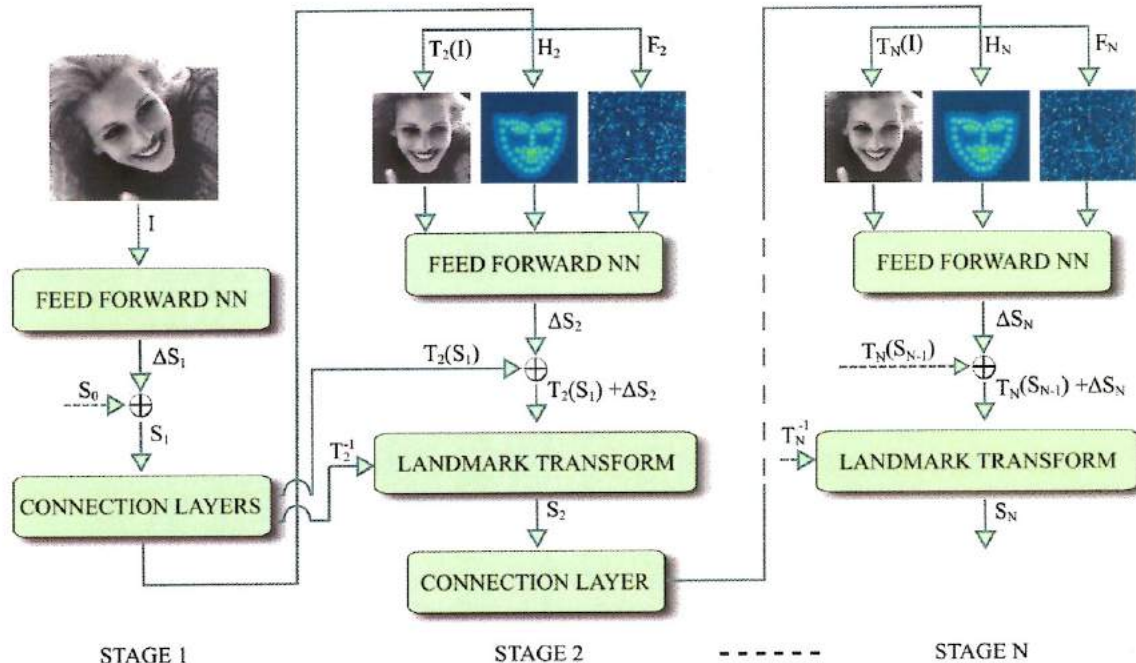
W ostatnich latach większość rozwiązań w zakresie analizy twarzy zdominowanych zostało przez sieci neuronowe. Sukces tych metod można przypisać w dużej mierze rozwinięciu aparatu matematycznego, dostępności olbrzymich zbiorów danych, jak również znacznemu zwiększeniu możliwości obliczeniowych współczesnych komputerów, poprzez wykorzystanie potencjału kart graficznych do obliczeń ogólnego przeznaczenia.

W pracy [H3] przedstawiony został algorytm *Deep Alignment Network*, który, podobnie jak *K-cluster Regression Forests with Weighted Splitting*, służy do lokalizacji punktów szczególnych twarzy. Metoda oparta jest na splotowych sieciach neuronowych i przebiega w sposób kaskadowy, gdzie kolejne stopnie kaskady prowadzą do zwiększenia precyzji lokalizacji punktów.

Każdy stopień kaskady składa się ze splotowej sieci neuronowej typu *feed-forward*, oraz szeregu modułów nazwanych warstwami łączącymi (*connection layers*). Sieć splotowa złożona jest z 14 warstw – 8 warstw splotu, 4 warstw typu *max-pooling* oraz dwóch warstw gęstych. Na wejściu sieci podawany jest obraz w skali szarości o wymiarach 112x112 pikseli. **Warstwy łączące, będące unikalną cechą tego algorytmu, służą do przekazywania informacji zgromadzonych w danym stopniu kaskady do kolejnych stopni kaskady**. Na wyjściu warstwy łączącej tworzone są trzy obrazy – mapa pozycji zlokalizowanych w danym stopniu punktów szczególnych (tzw. *heat map*), obraz twarzy znormalizowany do tzw. pozy kanonicznej oraz obraz reprezentujący wyjście pierwszej warstwy gęstej sieci neuronowej. Schemat proponowanego algorytmu przedstawiony jest na rysunku 4.

Za pomocą prezentowanej metody udało się uzyskać precyzję lokalizacji na poziomie **najlepszych istniejących metod, przewyższające także metodę KRFWS**. W konkursie na najlepszą metodę lokalizacji punktów szczególnych twarzy, który organizowany był na **warsztacie konferencji Computer Vision and Pattern Recognition (CVPR)**, będącej jedną z najważniejszych konferencji dotyczącej analizy obrazów na świecie, **w 2017 roku zajęła ona trzecie miejsce spośród dziesięciu konkurentów**. Wyniki te dowodzą, że skuteczność proponowanego algorytmu jest na poziomie najlepszych dostępnych obecnie metod.

Ze względu na fakt, że algorytm *Deep Alignment Network* zrealizowany jest z wykorzystaniem sieci neuronowych, do działania w czasie rzeczywistym wymaga karty graficznej o wysokiej wydajności.



Rysunek 4: Schemat algorytmu DAN [H3]. W każdym stopniu kaskady (*stage*) pozycja punktów szczególnych jest doprecyzowywana, zaczynając od początkowej estymacji S_0 . T określa transformację do pozycji kanonicznej w danym stopniu kaskady, H określa mapę pozycji zlokalizowanych punktów szczególnych, F określa wyjście pierwszej warstwy gęstej poprzedniej sieci typu *feed forward*.

Kody źródłowe algorytmu DAN, służące zarówno do jego uczenia jak i testowania metody, dostępne są publicznie w serwisie GitHub.

5.B. Algorytmy 3W modelowania i rekonstrukcji twarzy

Modelowanie i analiza obrazów trójwymiarowych twarzy stanowi wyzwanie, którego podejmuje się wiele ośrodków naukowych na świecie. Wzrost zainteresowania tym tematem wynika z dostępności tanich kamer głębi, takich jak *Microsoft Kinect*, czy *Intel Realsense*, a także większymi możliwościami obliczeniowymi współczesnych komputerów. W przedstawionym cyklu publikacji przebadany został temat analizy danych trójwymiarowych w celu konstrukcji deformowalnego modelu twarzy.

Powszechnie dostępne kamery głębi działają zazwyczaj na zasadzie stereo-pary lub poprzez analizę światła strukturalnego. Piksele obrazu, będącego rezultatem procesu akwizycji, określają odległość obiektów odpowiadających poszczególnym punktom do kamery. Pełną reprezentację trójwymiarowego obiektu można uzyskać poprzez ustawienie wielu kamer głębi w różnych punktach obserwacji tak, aby objąć całą strukturę obiektu. Aby odwzorować piksel głębi na pozycję trójwymiarową w układzie kamery rejestrującej, znane muszą być parametry wewnętrzne tej kamery (ang. *intrinsic camera parameters*). Aby uzyskać pełny, trójwymiarowy obraz obiektu, punkty otrzymane z kolejnych widoków muszą być przekształcone do wspólnego układu odniesienia. Przekształcenie to realizuje się z wykorzystaniem tzw. zewnętrznych parametrów kamery (ang. *extrinsic camera parameters*).

W publikacji [H4] przedstawiono efektywną metodę wyznaczania parametrów transformacji 3W za pomocą algorytmu *iterative closest points*, umożliwiającą rejestrację chmur punktów z wykorzystaniem obliczeń w architekturze kart graficznych. W proponowanym algorytmie, w celu zawężenia zbioru poszukiwań najbliższych sąsiadów, przestrzeń trójwymiarowa dzielona jest na segmenty o równych wielkościach. Wybrany sposób dekompozycji pozwala na efektywne wykorzystanie wielu procesorów dostępnych w kartach graficznych.



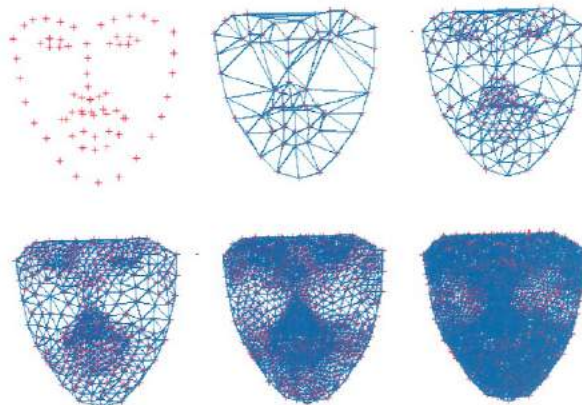
Rysunek 5: Przykład rejestracji modelu 3W z wykorzystaniem trzech kamer głębi Microsoft Kinect v2.

Doświadczenie zdobyte w ramach pracy nad publikacją [H4] wykorzystano do stworzenia deformowalnego, trójwymiarowego modelu twarzy, opisanego w artykule [H5]. Realizacja tego zadania zawiera się w trzech etapach: rejestracji trójwymiarowej bazy twarzy, wyznaczenia punktów charakterystycznych oraz wyznaczenia składowych elementów modelu.

Akwizycja danych trójwymiarowych zrealizowana została z wykorzystaniem trzech kamer głębi *Microsoft Kinect v2*. Urządzenia te, poza obrazem głębi, dostarczają także obraz RGB z mapą odwzorowania pozycji pikseli obrazu głębi w obrazie RGB. Kamery ustawione zostały na statywach z przodu oraz z boków obserwowanej osoby w odległości ok. 1 metra. Tył głowy nie był rejestrowany. Obrazy głębi były następnie odwzorowywane w punkty trójwymiarowe w układzie kamery, które następnie transformowane były do wspólnego układu odniesienia. Wartość koloru poszczególnych punktów przypisywana była na podstawie wartości odpowiadających im kolorów w obrazie RGB. Przykładowy skan zarejestrowany w przedstawiony powyżej sposób przedstawia rysunek 5.

Do dalszej analizy zarejestrowane zostały skany 26 osób, z których każda wykonywała określony zestaw 12 ekspresji twarzy (jak zamknięcie oka, uśmiech, smutek itp.). W rezultacie otrzymano 284 obrazy trójwymiarowe.

Aby możliwe było stworzenie 3W modelu aktywnego kształtu, zaproponowano autorską metodę określania korespondencji pomiędzy trójwymiarowymi punktami twarzy. W proponowanej metodzie, w pierwszym kroku, realizowana jest projekcja ortogonalna wszystkich punktów trójwymiarowych na płaszczyznę równoległą do powierzchni twarzy. W otrzymanym obrazie wykrywana jest twarz i lokalizowane są jej punkty szczególne (w tym przypadku za pomocą algorytmu KRFWS [H2]). Aby utworzyć „gęsty” model, który dla każdego obrazu charakteryzuje się tą samą liczbą odpowiadających sobie semantycznie wierzchołków, algorytm rozpoczyna działanie od zbioru „rzadkiego” - 68 wykrytych uprzednio punktów szczególnych. Pierwszym krokiem „zagęszczenia” modelu jest dwuwymiarowa triangulacja Delaunay’a na aktualnym zbiorze punktów. Dla każdego z powstałych trójkątów wyznaczany



Rysunek 6: Proces tworzenia gęstej siatki punktów charakterystycznych na podstawie bazowego zbioru 68 punktów szczególnych. Pierwszy obraz pierwszego rzędu określa pozycje wyznaczonych punktów charakterystycznych, natomiast drugi ilustruje efekt algorytmu trójkątowania Delaunaya. Na kolejnych rysunkach przedstawiono kolejne iteracje algorytmu dodawania nowych punktów wewnątrz znalezionych trójkątów.

jest nowy punkt określony przez średnią pozycję wierzchołków wchodzących w skład tego trójkąta. Łącząc ten punkt z trzema krawędziami tworzone są trzy nowe trójkąty. Następnie proces dodawania nowych wierzchołków jest powtarzany, tym razem na zbiorze rozszerzonym o dodane elementy. Proces iterowany jest do momentu uzyskania pożądanej gęstości modelu (rys. 6). Kolejność wierzchołków tworzących nowe trójkąty jest zapamiętywana i procedura może być powtarzana dla dowolnego zestawu 68 punktów szczególnych.

Każdemu wierzchołkowi jest przypisywana pozycja trójwymiarowa, określona jako pozycja najbliższego sąsiada w przestrzeni 2W, spośród punktów tworzących projekcję obrazu trójwymiarowego na płaszczyznę równoległą do płaszczyzny twarzy.

Procedura tworzenia gęstego modelu powtarzana jest dla każdego z 284 obrazów trójwymiarowych. Mając do dyspozycji zbiór gęstych modeli twarzy dla różnych ekspresji, w których liczba punktów jest zawsze ta sama, można przejść do etapu tworzenia modelu. Zastosowano tutaj klasyczne podejście, w którym modelowana twarz określona jest jako liniowa kombinacja punktów twarzy średniej S_0 i zestawu punktów twarzy własnych S_i . Wyznaczenie twarzy S_0 zrealizowane jest poprzez uśrednienie kształtów wszystkich dostępnych przykładów. Twarze własne S_i wyznaczone są w sposób automatyczny, np. poprzez metodę analizy składowych głównych lub poprzez odjęcie od kształtu S_0 uśrednionych zestawów przykładów, odzwierciedlających możliwe wariacje pożądanego kształtu. Rys. 7 przedstawia zestaw twarzy własnych uzyskanych na podstawie zbiorów punktów rejestrowanych dla różnych osób dla konkretnych ekspresji twarzy.

Wyniki eksperymentalne potwierdziły, że stworzony za pomocą powyższej metody model umożliwia reprezentację twarzy wyrażającej dowolne złożenie różnych ekspresji. Utworzone w ramach pracy stanowisko zostało wykorzystane w dalszych badaniach nad możliwością tworzenia trójwymiarowych rekonstrukcji całej sceny – w ramach projektu LiveScan3D, opisanego w osiągnięciach niewchodzących w skład niniejszego cyklu publikacji.



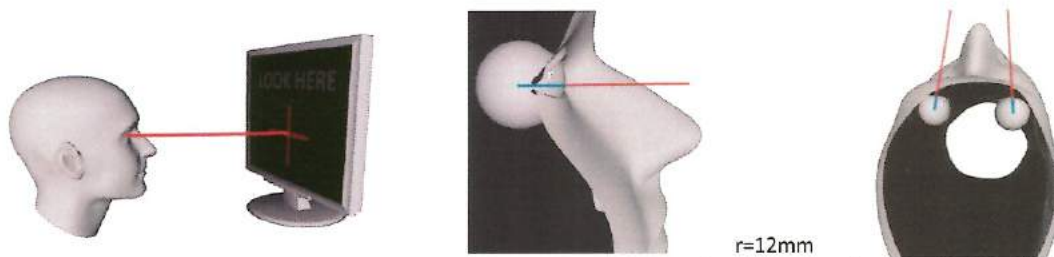
Rysunek 7: Zestaw punktów twarzy własnych dla różnych ekspresji twarzy. W kolejności od górnego lewego rogu do prawego dolnego rogu: „podniesienie brwi”, „podniesienie lewej brwi”, „zmarszczenie”, „zamknięcie lewego oka”, „zamknięcie prawego oka”, „krzyk”, „uśmiech”, „zaskoczenie”, „i”, „m”, „o”, „u”. W celu wizualizacji wyników do twarzy własnych dodana jest twarz średnia.

5.C. Zastosowania metod analizy twarzy

Lokalizacja punktu fiksacji wzroku w obrazach RGB jest tematem, dla którego, na chwilę obecną, nie istnieją rozwiązania o skuteczności umożliwiającej ich komercyjne zastosowanie. Istniejące systemy określania kierunku patrzenia zwykle opierają się na kamerach i oświetlaczach podczerwieni, wykorzystując zjawisko ciemnej lub jasnej źrenicy. W rzeczywistości wiele praktycznych zastosowań wyklucza wprowadzanie do stanowiska dodatkowych oświetlaczy i kamer. Co więcej, zastosowanie światła podczerwonego wyklucza także zastosowanie algorytmu w warunkach światła słonecznego. Z tego względu, w ramach opisanego tutaj osiągnięcia naukowego, opracowana oraz zaimplementowana została nowa metoda wykrywania punktu fiksacji wzroku w kamerach RGB. Wprowadzono **szereg kontrybucji do istniejącego stanu wiedzy**, w tym **unikalną metodę jednopunktowej kalibracji**, wydajny, **działający na poziomie state-of-the-art algorytm wykrywania źrenicy oka** oraz **algorytm śledzenia trójwymiarowej pozycji środka oka**.

Algorytmy wykrywania punktu fiksacji wzroku dzielą się zwykle na dwie kategorie. W pierwszej z nich, najpopularniejszej, jeśli chodzi o algorytmy opublikowane w postaci dostępnych aplikacji, kierunek patrzenia wyznaczany jest na podstawie pozycji oka oraz źrenicy w obrazie. Druga kategoria, do której należy także prezentowany tutaj system, opiera się na trójwymiarowym modelu geometrycznym. Punkt fiksacji określony jest poprzez punkt przecięcia prostej przechodzącej przez trójwymiarową pozycję środka oka oraz źrenicy z płaszczyzną ekranu.

Pierwszym krokiem proponowanego w patencie [H6] algorytmu jest kalibracja, która polega na określeniu początkowej, trójwymiarowej pozycji środka oka. Użytkownik proszony jest o skupienie wzroku na określonym punkcie ekranu (rysunek 8, lewa strona). Następnie realizowana jest detekcja twarzy i jej punktów charakterystycznych, w tym źrenicy oka.



Rysunek 8: Ilustracja procesu kalibracji. Użytkownik obserwuje określony punkt na ekranie (lewa strona obrazu). Środek oka określony jest jako punkt leżący na prostej utworzonej przez punkt obserwacji i środek źrenicy (środek i prawa strona obrazu). Zakładamy, że punkt ten znajduje się w odległości $r=12\text{mm}$ od punktu przecięcia tej prostej z powierzchnią źrenicy w głąb oka.

Zakładając, że znane są parametry kamery, oraz przyjmując, że odległość między oczami u dorosłego człowieka wynosi ok. 6cm, można określić odległość oczu od obserwowanego punktu. W przybliżeniu środek oka leży na prostej określonej przez punkt obserwacji i środek źrenicy oka, w odległości $r=6\text{mm}$ od punktu przecięcia z powierzchnią źrenicy w głąb oka. Należy tutaj zauważyć, że odległości te (6cm i 6mm) mogą nieznacznie różnić się w zależności od osoby, jednak nasze eksperymenty wykazały, że uwzględnienie dokładnych wartości nie wpływa istotnie na dokładność śledzenia. W przyszłych badaniach można także uwzględnić, że oko nie jest idealnie kuliste - średnica pionowa rogówki wynosi ok. 11mm, natomiast pozioma ok. 12mm.

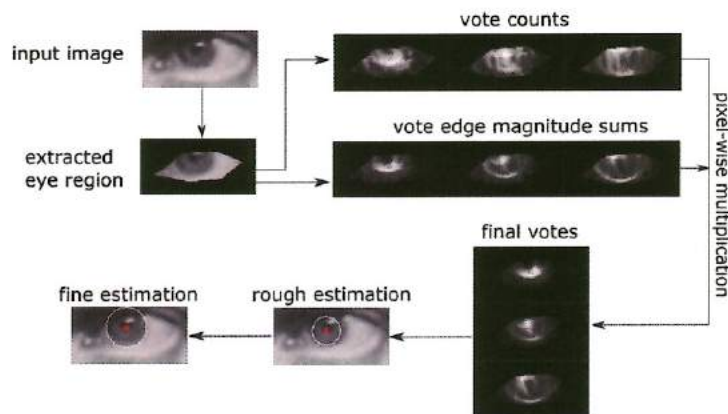
Wyznaczony w procesie kalibracji środek oka jest śledzony w kolejnych klatkach obrazu. Można założyć, że translacja i rotacja oka bezpośrednio odpowiada rotacji i translacji całej głowy. Z tego względu zastosowano metodę śledzenia trójwymiarowych parametrów pozy i przesunięcia głowy w wykorzystaniu algorytmu przepływu optycznego, gdzie twarz reprezentowana jest przez cylinder lub szczegółowy model trójwymiarowy. Więcej szczegółów na temat tego algorytmu znajduje się w [H6].

Algorytm wykrywania punktu fiksacji wzroku realizowany jest w czasie rzeczywistym. W każdym kolejnym obrazie śledzony jest trójwymiarowy środek oka oraz lokalizowana jest dwuwymiarowa pozycja źrenicy. Pozycję trójwymiarową źrenicy można z łatwością określić dysponując wewnętrznymi parametrami kamery oraz przyjmując, że leży ona na kuli zaczepionej w środku oka, której promień wynosi ok. 6mm. Punkt fiksacji w każdej klatce określony jest poprzez wyznaczenie punktu przecięcia prostej przechodzącej przez punkt środka oka i środka źrenicy, z obserwowaną płaszczyzną (np. ekranem).

Proponowany algorytm, w przeciwieństwie do wielu istniejących metod, umożliwia na wykrywanie punktu fiksacji z zachowaniem naturalnych ruchów głowy, tj. bez konieczności zachowania nieruchomej pozycji. Istotną wadą proponowanego algorytmu jest akumulacja błędów dryftu wynikającego z zastosowania metody przepływu optycznego. Z tego względu konieczna jest okresowa reinicjalizacja modelu (algorytm reinicjalizacji został także zaproponowany w patencie [H6]). Błąd algorytmu oszacowano na poziomie poniżej 2° dla głowy nieruchomej i poniżej 4° dla naturalnych ruchów głowy. Błąd ten, niestety, zwiększa się z czasem. Problem ten był analizowany w kolejnych publikacjach, nie ujętych w opisie niniejszego osiągnięcia naukowego.



Rysunek 9: Osoba badana z wykorzystaniem okulografu *JazzNovo* firmy *Ober Consulting* [H7].



Rysunek 10: Kroki algorytmu wyznaczania środka i promienia tęczówki [H7].

W artykule [H7] przedstawiony został system służący do śledzenia źrenicy oka z wykorzystaniem standardowej kamery RGB. Celem stworzonej w ramach badań aplikacji jest realizacja testu, na podstawie którego lekarz specjalista może ocenić stan zaawansowania choroby Parkinsona u pacjentów. Istniejące sprzętowe, komercyjne okulografy, jak np. okulograf firmy *Ober Consulting*, będący referencją dla proponowanego systemu, są często niewygodne dla pacjenta i wykluczają wykonanie badania samodzielnie w warunkach domowych (rys. 9). Okulografy charakteryzują się natomiast wysoką dokładnością oraz wysoką częstotliwością próbkowania. Opracowany i przedstawiony w [H7] system wideo-okulografii, opierający się na standardowej, łatwo dostępnej kamerze, umożliwia pacjentom samodzielne wykonanie badania, ograniczając rolę lekarza jedynie do weryfikacji otrzymanych wyników. W realizowanym zastosowaniu ocenie podlegała:

- koordynacja ruchów nadążnych – pacjent obserwuje płynnie przesuwaną się obiekt na ekranie,
- koordynacja ruchów sakadowych – pacjent obserwuje obiekt przemieszczający się w sposób skokowy,
- koordynacja ruchów antysakadowych – pacjent ma za zadanie patrzeć w kierunku przeciwnym niż skokowo przemieszczający się obiekt.

Proponowany w pracy [H7] algorytm, podobnie jak algorytm wykrywania punktu fiksacji oka, rozpoczynany jest od detekcji twarzy. Następnie twarz i jej punkty szczególne są śledzone w kolejnych klatkach obrazu hybrydowym algorytmem łączącym klasyczną lokalizację punktów (takich jak np. KRFWS) i śledzenie trójwymiarowego modelu twarzy. Następnie w obszarze oczu wykrywany jest środek tęczówki.

Lokalizacja tęczówki przeprowadzana jest w obszarze określonym przez punkty charakterystyczne znajdujące się na obrysie oka. W pierwszym kroku przeprowadzane jest głosowanie na pozycję środka tęczówki za pomocą algorytmu *symmetry radial transform*. Działanie tego algorytmu przypomina transformatę Hougha dla okręgów (ang. *Circle Hough Transform*), z tą różnicą, że głosowanie przebiega jedynie w linii prostopadłej do krawędzi danego piksela, w kierunku od obszaru jasnego do obszaru ciemnego. Liczba głosów oddanych na dany piksel jest następnie mnożona przez sumę siły krawędzi biorących udział w głosowaniu na dany piksel. Operacja ta powtarzana jest dla wszystkich możliwych skal źrenicy oka. Z uzyskanych wartości szacowana jest wstępna pozycja i średnica tęczówki oka. W ostatnim kroku, pozycja ta jest doprecyzowywana z wykorzystaniem całkowo-różniczkowego operatora Daugmanna. Większość operacji wykonywanych jest w dokładności podpikselowej z wykorzystaniem interpolacji dwuliniowej. **Zaproponowany algorytm wykrywania środka tęczówki charakteryzuje się bardzo wysoką precyzją, przewyższającą metody dostępne w momencie publikacji.** Poszczególne etapy algorytmu zilustrowane są na rys. 10. Ostateczna pozycja źrenicy określona jest w stosunku do pozycji środka oka, a jej wyniki znormalizowane są do przedziału $\langle -1, 1 \rangle$. Ze względu na potrzeby aplikacji analizowany jest jedynie ruch w kierunku poziomym. Dokonano także porównania prezentowanego systemu z referencyjnym, komercyjnym okulografem *JazzNovo*, które wykazały średnią korelację pomiędzy wynikami obu rozwiązań na danych testowych na poziomie 0,864. Większość niedopasowań zaobserwowano u pacjentów, mających mocno przymknięte oczy, co powodowało, że zarówno proponowany system jak i system referencyjny charakteryzował się wysokim poziomem szumu. Należy także zauważyć, że zadaniem systemu nie było określanie kierunku patrzenia, a raczej szybkość reakcji na określone bodźce wzrokowe.

Ograniczeniem systemu jest częstotliwość próbkowania kamery wideo. Z tego względu do badań wykorzystana była kamera umożliwiająca rejestrację obrazów z częstotliwością 60Hz (*Logitech C922*).

Algorytm lokalizacji tęczówki, wraz z danymi niezbędnymi do porównania do innych algorytmów, udostępniony został w serwisie GitHub jako projekt typu open-source (www.github.com/JacekNaruniec/IrisLocalization).

W pracy [H8] przedstawiono problem identyfikacji osoby na podstawie obrazu ust. Jest to temat znacznie słabiej poznany niż identyfikacja osoby na podstawie obrazu całej twarzy, który cieszy się w obecnej chwili niezwykle popularnością. Zdarza się jednak, szczególnie w przypadku pracy policyjnej, np. w czasie identyfikacji sprawców przestępstw, że osoby rozpoznawane mają zakrytą większość część twarzy, ale usta są widoczne. W takich przypadkach, mimo niższej skuteczności od algorytmów opartych na pełnym obrazie,

algorytmy takie mogą okazać się niezwykle przydatne. W pracy [H8] zaproponowano metodę wykorzystującą probabilistyczne sieci neuronowe (*Probabilistic Neural Networks*) w połączeniu z optymalizacją rojów cząsteczek (*Particle Swarm Optimization*) do identyfikacji osób na podstawie obrazów ust. W algorytmie tym wykorzystano przedstawiony w ramach niniejszego opracowania algorytm KRFWS do lokalizacji punktów szczególnych ust. Jest to element niezwykle istotny, gdyż poprawna lokalizacja tych punktów jest kluczowa dla poprawnego wyznaczenia współczynników służących do budowy deskryptora obrazu. W proponowanym algorytmie wykorzystany został zestaw cech statystycznych, służący jako wejście do algorytmu opartego na probabilistycznych sieciach neuronowych i optymalizacji rojów cząsteczek. Uzyskane wyniki są obiecujące, jednak w chwili obecnej testy dotyczą jedynie twarzy frontalnych, co jest pewnym ograniczeniem proponowanej metody, szczególnie w stosunku do istniejących obecnie metod rozpoznawania twarzy.

6. Podsumowanie osiągnięcia wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.).

W mojej ocenie, do najważniejszych osiągnięć opisanych w wymienionych publikacjach, stanowiących osiągnięcie w rozumieniu art. 16 ust. 2 ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 2016 r. poz. 882 ze zm. w Dz. U. z 2016 r. poz. 1311.) należą:

- **Artykuł H1:**
 - Opracowanie nowego algorytmu wykrywania punktów szczególnych twarzy, będący rozwinięciem moich badań prowadzonych w ramach pracy doktorskiej, charakteryzujący się precyzją przewyższającą wiele najlepszych metod istniejących w momencie publikacji.
 - Opracowanie nowego algorytmu ekstrakcji cech obrazu – *discrete area filters*.
 - Opracowanie algorytmu umożliwiającego jednoczesną lokalizację 15 punktów szczególnych i detekcję twarzy.

- **Artykuł H2:**
 - Wykazanie, że zastosowanie drzew losowych i maszyny wektorów nośnych (*Support Vector Machine*) w problemie lokalizacji punktów szczególnych umożliwia zwiększenie skuteczności lokalizacji charakterystycznych punktów twarzy w stosunku do metody opartej na regresji dwustopniowej.
 - Opracowanie nowego mechanizmu tworzenia drzew decyzyjnych opartego na wagowaniu przykładów oraz klasyfikacji za pomocą maszyny wektorów nośnych.
 - Opracowanie algorytmu lokalizacji punktów szczególnych przewyższającego skutecznością lokalizacji większość istniejących w momencie publikacji najlepszych metod.

- **Artykuł H3:**
 - Opracowanie metody przekazywania informacji pomiędzy stopniami kaskady sieci neuronowej za pomocą warstw łączących.

- *Opracowanie algorytmu lokalizacji punktów szczególnych przewyższającego skutecznością lokalizacji większość najlepszych metod istniejących w momencie publikacji.*
- **Artykuł H4:** *Opracowanie algorytmu dekompozycji przestrzeni punktów i opracowanie rozwiązań umożliwiających implementację algorytmu ICP w architekturze kart graficznych.*
- **Artykuł H5:**
 - *Opracowanie metody „zagęszczania” 68 punktowego modelu twarzy.*
 - *Opracowanie deformowalnego, trójwymiarowego modelu twarzy.*
- **Patent H6:**
 - *Opracowanie metody jednopunktowej kalibracji w zadaniu śledzenia punktu fiksacji oka.*
 - *Opracowanie efektywnej metody śledzenia pozycji środka oka.*
 - *Opracowanie metody wykrywania punktu fiksacji oka z wykorzystaniem geometrycznego modelu twarzy.*
- **Artykuł H7:**
 - *Opracowanie skutecznego algorytmu lokalizacji tęczówki oka, który charakteryzuje się precyzją przewyższającą wiele najlepszych metod istniejących w momencie publikacji.*
 - *Opracowanie systemu analizy twarzy umożliwiającego zastąpienie sprzętowego okulografu algorytmem wizyjnym, wykorzystanego w badaniu oceny postępów choroby Parkinsona.*
- **Artykuł H8:** *Opracowanie algorytmu identyfikacji osoby na podstawie obrazu ust, z wykorzystaniem probabilistycznej sieci neuronowej i optymalizacji roju cząsteczek.*



Rysunek 11: Przykład rekonstrukcji trójwymiarowej za pomocą aplikacji LiveScan3D.

7. Pozostałe osiągnięcia naukowo-badawcze i popularyzatorskie.

Moje zainteresowania badawcze, poza przedstawionymi w ramach powyższego osiągnięcia naukowego, dotyczą m.in. tematyki rekonstrukcji trójwymiarowej oraz rzeczywistości rozszerzonej. Opublikowany na konferencji naukowej 3DV artykuł pt.: „*Livescan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors*” przedstawia system rekonstrukcji trójwymiarowej nie tylko twarzy, ale całej sylwetki, czy też sceny. Oprócz rekonstrukcji, system pozwala na przesyłanie informacji trójwymiarowej do urządzeń rozszerzonej rzeczywistości, takich jak *Microsoft HoloLens*. Upubliczniona w serwisie GitHub aplikacja i kody źródłowe LiveScan3D cieszą dużą popularnością (na dzień dzisiejszy 227 gwiazdek i 91 kopii). System LiveScan3D był również wdrożony w ramach dwóch spektakli teatralnych, gdzie „na żywo” wykonywana i wizualizowana była trójwymiarowa rekonstrukcja tancerzy oraz publiczności. System jest aktywnie rozwijany, w chwili obecnej prace koncentrują się na rekonstrukcji powierzchni tworzonych przez chmury punktów (rys. 11).

Kolejnym podejmowanym przeze mnie wyzwaniem badawczym i implementacyjnym jest rozpoznawanie twarzy. W ostatnich latach, w ramach projektu BioWiz zrealizowałem algorytm umożliwiający na identyfikację osób z bardzo wysoką skutecznością, dorównująca najlepszym, dostępnym obecnie algorytmom. Proponowana przeze mnie metoda oparta jest na najnowszych architekturach splotowych sieci neuronowych. Stworzona w ramach zadań projektowych biblioteka rozpoznawania twarzy jest obecnie wdrażana w systemach Komendy Głównej Policji.

W ramach mojej pracy dydaktycznej prowadziłem 28 prac dyplomowych, które zakończyły się obroną z wynikiem pozytywnym. Dwukrotnie, w latach 2014 i 2017, zostałem przez studentów wybrany na najlepszego wykładowcę Wydziału Elektroniki i Technik Informatycznych w konkursie „Złota Kreda”. Byłem prowadzącym oraz przygotowałem wykłady i laboratoria czterech przedmiotów: „Analiza Semantyczna Obrazu i Dźwięku”, „Programowanie Obiektowe”, „Analiza i Indeksowanie Multimediów” oraz „Szybkie Masywnie Równoległe Metody Obliczeniowe”. W 2014 roku prowadziłem także specjalny wykład szkoleniowy dla pracowników firmy Samsung R&D pt.: „Kluczowe algorytmy przetwarzania obrazu”.