



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

DZIEDZINA NAUK INŻYNIERYJNO-TECHNICZNYCH

DYSCYPLINA AUTOMATYKA, ELEKTRONIKA, ELEKTROTECHNIKA I TECHNOLOGIE
KOSMICZNE

AUTOREFERAT ROZPRAWY DOKTORSKIEJ

Nowe metody rozpoznawania gestów 3D bez użycia
kamer, w aplikacjach w branży motoryzacyjnej

Autor: mgr inż. Piotr Reczek

Promotor rozprawy: dr hab. inż. Andrzej Młyniec, prof. AGH

Promotor pomocniczy: dr hab. inż. Andrzej Wetula, prof. AGH

Praca wykonana: Akademia Górniczo-Hutnicza im. Stanisława
Staszica w Krakowie

Wydział Inżynierii Mechanicznej i Robotyki

Katedra Robotyki i Mechatroniki

Kraków, 2023

1. Streszczenie

Niniejsza praca przedstawia nowe metody rozpoznawania gestów 3D bez użycia kamer, w aplikacjach w branży motoryzacyjnej. W ramach realizacji pracy, dokonano przeglądu literatury oraz istniejących rozwiązań, które mogłyby zostać zastosowane do rozpoznawania gestów w środowisku wnętrza pojazdu. Następnie przeanalizowano istniejące systemy w następujących kategoriach: Systemy wizyjne, Systemy wykorzystujące elektrody pojemnościowe, Systemy oparte o czujniki podczerwieni, Systemy radarowe, Systemy ultradźwiękowe, Systemy zlokalizowane na ręce. Każdą z zaproponowanych technologii poddano krytycznej analizie z perspektywy wymagań dla systemów do rozpoznawania gestów z potencjalnym zastosowaniem w branży samochodowej. Dodatkowo, każdą z nich przeanalizowano pod kątem wykorzystania do utworzenia nowego rozwiązania wykluczając z założenia systemy wizyjne. Na zakończenie przeglądu literatury omówiono fuzję danych oraz klasyfikację metod wykorzystywanych do jej przeprowadzania. W rozdziałach „System do rozpoznawania gestów oparty o czujniki pojemnościowe” oraz „System do rozpoznawania gestów oparty o radar”, przedstawiono autorskie rozwiązania, w tym zastosowaną architekturę sprzętową, opracowane algorytmy oraz sposoby działania obu systemów. Rozwiązania te zostały stworzone w ramach realizacji projektu wdrożeniowego w firmie Merit. W rozdziale traktującym o systemie pojemnościowym zaproponowano również sposoby rozszerzania danych uczących dla przebiegów sygnału pojemnościowego z poszczególnych elektrod. W końcowej części rozdziału opisano proces konwersji modelu sieci neuronowej oraz uruchomienia go na mikrokontrolerze. W kolejnym podrozdziale opisana została implementacja systemu działającego w oparciu o czujniki pojemnościowe, który został zamontowany w samochodzie testowym, co stanowiło wdrożeniowy charakter pracy. Na zakończenie rozdziału, przedstawiono nowatorską metodę automatyzacji akwizycji danych uczących dla potrzeb procesu treningu sieci neuronowej przy użyciu 4 osiowego manipulatora, który zastąpił ludzi wykonujących gesty. W rozdziale dotyczącym systemu używającego radaru fali ciągłej z modulacją częstotliwościową, oprócz opisu konstrukcji radaru oraz jego sprzętowej konfiguracji, przedstawiono zastosowane metody przetwarzania danych otrzymywanych bezpośrednio z radaru do postaci cech, które stanowiły wejście do sieci neuronowej klasyfikującej gesty. W ostatnim rozdziale zostały przedstawione wyniki skuteczności rozpoznawania gestów dla obu opisanych rozwiązań, a także po dokonaniu fuzji danych pochodzących z każdego z nich. Sieć neuronowa wykorzystana do syntezy danych, przyjmowała rozszerzony wektor wejściowy obejmujący 10 cech w postaci 5 sygnałów z czujnika pojemnościowego, 2 przebiegów określających prędkość obiektu względem anten radaru, 2 przebiegów informujących o odległości wykrywanego obiektu od każdej z anten oraz położenia kątownego wykrywanego obiektu w płaszczyźnie wertykalnej radaru. Przeprowadzone eksperymenty

wykazały, że modele sieci neuronowej rozpoznające zarówno 5 oraz 7 gestów, dzięki zastosowaniu fuzji danych, osiągnęły ponad 97 procentową skuteczność ich rozpoznawania, podczas gdy sieci neuronowe wykorzystujące dane tylko z jednego typu sensora (elektrod pojemnościowych i radaru) odpowiednio dla 5 gestów 93% i 94,2% oraz dla 7 gestów 87,8% i 92,5%. Otrzymane rezultaty dowiodły postawionej tezie, że fuzja danych z obu systemów spowoduje wzrost skuteczności wykrywanych gestów. W ostatnim rozdziale przedstawiono obszary, których udoskonalenie powinno zostać rozważone pod kątem dalszego rozwoju przedstawionych systemów, zarówno w kontekście wykorzystania osobno jak i w kontekście ich fuzji.

2. Wstęp i motywacja

Ilość funkcji dostępnych w pojazdach wzrasta systematycznie niemal od początku istnienia motoryzacji. Liczba funkcji dostępnych we współczesnych samochodach, często powoduje zagubienie użytkowników oraz wymaga czasochłonnego adaptowania się do obsługi pojazdu. Stanowi to wyzwanie dla producentów sektora motoryzacyjnego, które polega na poszukiwaniu i rozwoju dogodnych sposobów komunikowania się z samochodem. Nowe interfejsy powinny ułatwiać kontrolę dostępnych systemów, działać w sposób niezawodny oraz intuicyjny. Wymogi te skutkują licznymi poszukiwaniami nowych rozwiązań przez przemysł oraz środowisko akademickie.

Tytuł rozprawy wykluczający z zakresu rozważanych rozwiązań systemy wizyjne, motywowany był poszukiwaniami rozwiązań o mniejszym stopniu skomplikowania, a w efekcie o niższym koszcie rozwoju oraz produkcji. Jest to motywowane zapewnieniem możliwości wykorzystywania interfejsu w postaci sterowania gestami w szerszej liczbie modeli samochodów, nie tylko luksusowych. Powyższe uwarunkowania miały bezpośrednie przełożenie na przeprowadzoną analizę technologii możliwych do zastosowania dla realizacji zadania rozpoznawania gestów, a także na kierunek badań i rozwoju systemów przeprowadzonej w części praktycznej doktoratu.

3. Cel i teza pracy

Cel główny pracy, stanowiło opracowanie nowego systemu do rozpoznawania gestów, bazującego na fuzji systemów pojemnościowego i radarowego do wykorzystania we wnętrzu pojazdu samochodowego. Dodatkowo zrealizowano cele dodatkowe:

- Opracowanie krytycznej analizy dostępnych technologii, które mogą zostać wykorzystane do rozpoznawania gestów w środowisku wnętrza pojazdu
- Wybór istniejących rozwiązań oraz poszukiwanie nowych, które pozwolą uzyskać najlepsze rezultaty przy zachowaniu efektywności kosztowej
- Budowa prototypów systemów rozpoznających gesty oraz przetestowanie ich w środowisku wnętrza pojazdu
- Rozwój wiedzy w zakresie systemów rozpoznających gesty
- Rozwój wiedzy w zakresie uczenia maszynowego oraz ich implementacji na systemach wbudowanych
- Zbadanie potencjału wykorzystania fuzji danych z dwóch systemów w celu poprawy skuteczności rozpoznawania gestów
- Sprawdzenie wpływu zastosowania fuzji danych z dwóch systemów na liczbę rozpoznawanych gestów

Tezy niniejszej pracy stanowiły:

“Fuzja danych z dwóch systemów pozwoli na uzyskanie lepszych rezultatów niż w przypadku wykorzystania każdego z nich osobno. Dotyczy to głównie skuteczności wykrywanych gestów, co pozwoliłoby na większą niezawodność systemu pozwalającą na zastosowanie w pojeździe. “

“Fuzja danych pozwoli na osiągnięcie lepszych rezultatów w przypadku większej liczby gestów. “

W celu potwierdzenia powyższych też przeprowadzono eksperymentalne badania mające na celu sprawdzenie słuszności postawionych założeń. Do analiz wykorzystano dwa autorskie systemy, które również przygotowano w ramach realizacji niniejszego doktoratu – system pojemnościowy oraz radarowy. Dla obu systemów algorytm rozpoznający gesty stanowiły odpowiednio dobrane sieci neuronowe. W pierwszym etapie przeprowadzono proces uczenia maszynowego dla każdego z systemów osobno, a następnie po ich połączeniu z wykorzystaniem fuzji danych w postaci rozszerzonego modelu sieci neuronowej.

4. Struktura i zakres rozprawy doktorskiej

Rozprawa doktorska składa się z 8 rozdziałów obejmujących następujące 4 zasadnicze zagadnienia:

- Analizę oraz krytyczną ocenę istniejących rozwiązań wykorzystywanych do rozpoznawania gestów
- Przedstawienie wykonanego systemu rozpoznającego gesty opartego o elektrody pojemnościowe
- Przedstawienie wykonanego systemu rozpoznającego gesty wykorzystującego radar
- Opis oraz wyniki eksperymentu polegającego na zastosowaniu fuzji danych do rozpoznawania gestów

Zakres rozprawy doktorskiej obejmuje:

- Wprowadzenie do tematyki pracy (Rozdział 1)
- Przegląd oraz krytyczną analizę istniejących rozwiązań z wyszczególnieniem systemów wizyjnych, pojemnościowych, podczerwieni, radarowych, ultradźwiękowych oraz zlokalizowanych na ręce (Rozdział 2)
- Omówienie oraz klasyfikację metod fuzji danych (Rozdział 2)
- Podsumowanie stanu wiedzy oraz postawienie tez (Rozdział 3)
- Opis autorskiego systemu rozpoznającego gesty wykorzystującego elektrody pojemnościowe, z uwzględnieniem testów w środowisku pojazdu (Rozdział 4)
- Opis procesu konwersji sieci neuronowej w kontekście uruchomienia algorytmu z poziomu mikrokontrolera (Rozdział 4)
- Nową metodę automatyzacji akwizycji danych dla potrzeb uczenia maszynowego wykorzystującą 4-osiowy manipulator z układem symulującym impedancje ludzkiej ręki (Rozdział 4)
- Opis systemu rozpoznającego gesty wykorzystującego radar z zastosowaniem autorskiego sposobu ekstrakcji danych wejściowych (Rozdział 5)
- Przedstawienie procesu wyznaczenia parametrów pracy radaru w kontekście postawionych wymagań (Rozdział 5)
- Przedstawienie przeprowadzonego eksperymentu porównującego efekty zastosowania fuzji danych do rozpoznawania gestów z systemu radarowego oraz pojemnościowego (Rozdział 6)
- Porównanie wyników skuteczności rozpoznawania gestów dla obu systemów osobno oraz z uwzględnieniem fuzji (Rozdział 6)
- Podsumowanie osiągniętych wyników oraz omówienie dalszych możliwości rozwoju (Rozdział 7)
- Spis literatury (Rozdział 8)

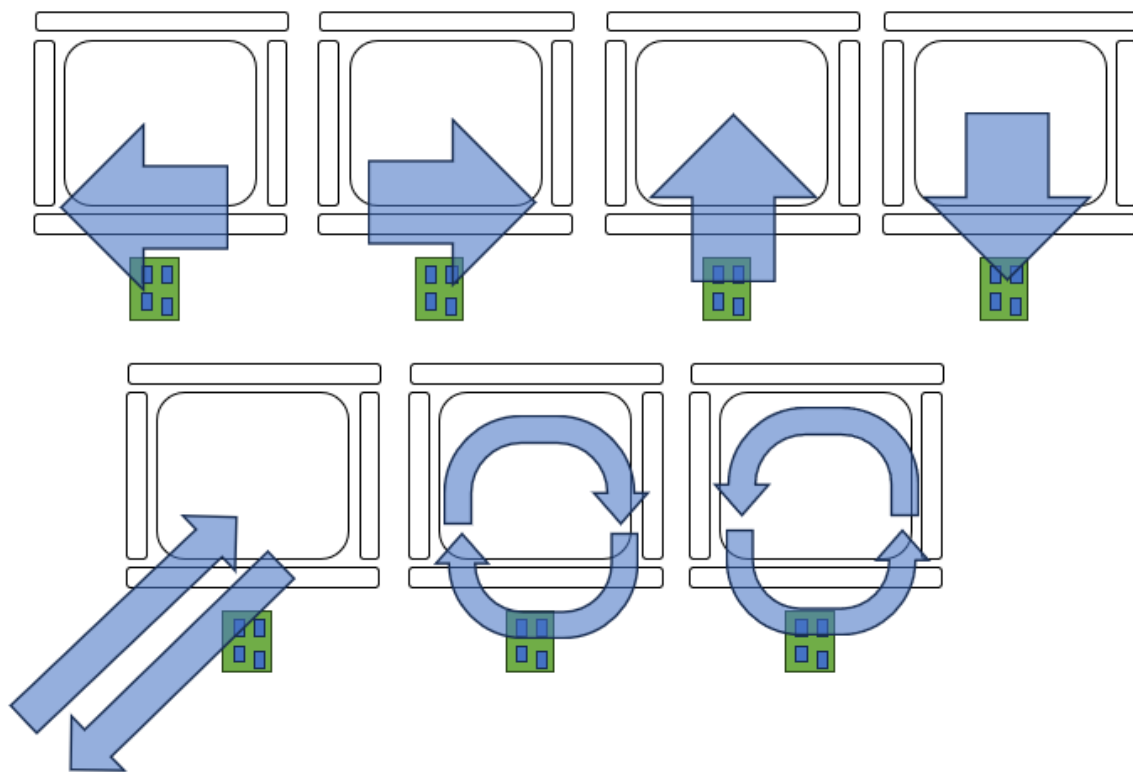
Rezultaty analiz oraz eksperymentów mogą stanowić syntetyczne źródło wiedzy dla inżynierów podejmujących tematykę rozpoznawania gestów. Część analityczna przedstawia porównanie dostępnych technologii, które mogą zostać wykorzystane przy rozwoju podobnych systemów również w przypadku innych aplikacji. Zaproponowany system automatyzacji akwizycji danych, przedstawia perspektywę znaczącego przyspieszenia procesów uczenia oraz testowania tego typu systemów. Przedstawione wnioski w wynikające z przeprowadzonej fuzji danych pokazują użyteczność wykorzystania danych z dwóch systemów pracujących w innych technologiach.

5. Podsumowanie oraz ważniejsze wyniki i wnioski

Efektom rozprawy jest:

- System pojemnościowy wraz z algorytmem rozpoznawania gestów opartym o sieć neuronową zaimplementowany w postaci systemu wbudowanego umiejscowionego w podłokietniku samochodu oraz zintegrowany z panelem centralnym pojazdu
- Eksperymentalna analiza wpływu na amplitudę mierzonego sygnału dla elektrod pojemnościowych dla konfiguracji materiałów osłonowych wykorzystywanych do zabudowania i zamaskowania elektrod.
- Analiza wpływu rozszerzania zestawu danych wykorzystywanych do uczenia modelu poprzez przeskalowywanie sygnału w osi czasu, wygładzanie przebiegu oraz dodanie zakłóceń.
- Przedstawienie algorytmów wykorzystanych do procesu trenowania modelu
- Proces konwersji oraz uruchomienia modelu sieci neuronowej z poziomu mikrokontrolera
- Weryfikacja koncepcji wykorzystania danych generowanych z użyciem manipulatora 4-osiowego pozwalającej na automatyzację procesu akwizycji danych
- System radarowy, gdzie zastosowano autorską kompilację algorytmów przetwarzających dane dla potrzeb sieci neuronowej
- Określenie parametrów pracy radaru optymalnych dla aplikacji rozpoznawania gestów
- Analiza możliwych rozwiązań z zakresu fuzji danych
- Przeprowadzenie eksperymentu weryfikującego tezy doktoratu, polegające na sprawdzeniu wzrostu skuteczności oraz liczby rozpoznawanych gestów z wykorzystaniem fuzji danych

Głównym eksperymentem przeprowadzonym na potrzeby sprawdzenia prawdziwości tez postawionych w rozprawie, było przeprowadzenie uczenia maszynowego dla każdego z systemów osobno, oraz wspólnie z wykorzystaniem fuzji danych. Na potrzeby przeprowadzenia eksperymentu zebrano 4000 gestów od 15 osób w wieku 25-65 lat. Ostatecznie do procesu uczenia modeli sieci neuronowych przeznaczonych do rozpoznawania gestów, wykorzystano zestaw danych zawierający 3284 gesty. Było to spowodowane niższą czułością systemu pojemnościowego, co skutkowało brakiem wykrycia rozpoczęcia części gestów przez ten system. Poniżej przedstawiono schemat obrazujący charakter wykonywanych gestów.



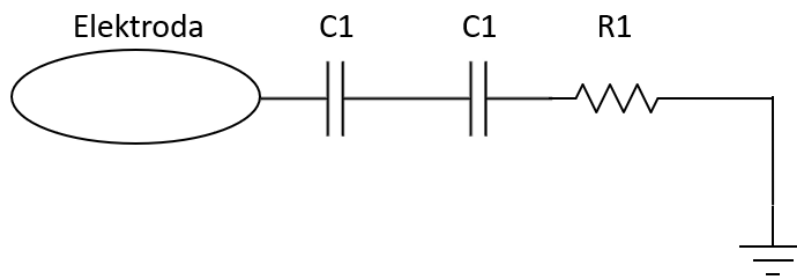
Rysunek 1. Rysunek przedstawiający wykonywane gesty w obszarze roboczym systemów. Kolejno: przesunięcie w lewo (swipe left), przesunięcie w prawo (swipe right), przesunięcie w górę (swipe up), przesunięcie w dół (swipe down), zbliżenie (tap), kołowy ruch zgodnie ze wskazówkami zegara (rotation clockwise), Kołowy przeciwnie do ruchu wskazówek zegara (rotation anticlockwise)

Dla wszystkich trzech systemów algorytm rozpoznawania gestów stanowiły sieci neuronowe o architekturze Inception Time [1]. Fuzja danych została przez stworzenie nowego modelu sieci neuronowej o rozmiarze wektora wejściowego uwzględniającym rozmiar danych wejściowych z obu systemów. W poniższej tabeli przedstawiono otrzymane rezultaty.

Tabela 1. Skuteczność rozpoznawania gestów dla poszczególnych modeli sieci neuronowej

5 Gestów			7 Gestów		
System radarowy	System pojemnościowy	System radarowy i pojemnościowy	System radarowy	System pojemnościowy	System radarowy i pojemnościowy
93.0%	94.2%	97.7%	87.8%	92.5%	97.1%

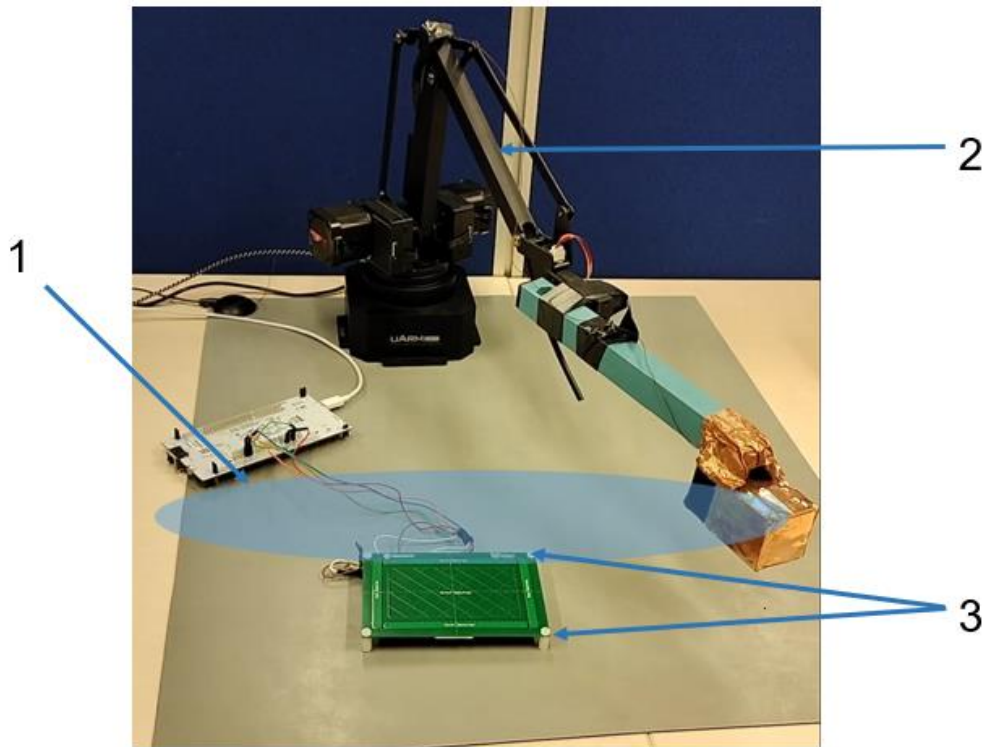
Kolejnym istotnym eksperymentem było udowodnienie, że możliwe jest wykorzystanie do procesu uczenia sieci neuronowej rozpoznającej gesty wykorzystującej elektrody pojemnościowe, danych pochodzących z gestów wykonywanych automatycznie. Automatyzacja polegała na generowaniu ruchów dedykowanego efektora za pomocą 4-osiowego manipulatora. Efektor miał za zadanie symulować parametry elektryczne ludzkiej ręki. W tym celu zasymulowano jej impedancję zgodnie z poniższym rysunkiem.



Rysunek 2. Układ elektroniczny symulujący impedancję ludzkiej ręki ($R1 = 1500\Omega$, $C1 = 200nF$)[2]

Parametry ruchu – współrzędne oraz prędkość, zostały podane procesowi randomizacji, dzięki czemu uzyskano rozrzut symulujący różnorodność danych spotykaną w zestawach gestów uzyskiwanych od

rzeczywistych użytkowników. Poniższy rysunek prezentuje stanowisko, na którym przeprowadzono eksperyment.



Rysunek 2. Stanowisko pomiarowe przedstawiające schemat ideowy eksperymentu. 1 – Chmura punktów w której porusza się efektor, 2 – Manipulator, 3 – Obszar roboczy elektrod.

Po przeprowadzeniu procesu zbierania danych oraz uczenia modelu wykorzystywanego do rozpoznawania gestów przez system pojemnościowy przeprowadzono test skuteczności rozpoznawania gestów. Do tego celu wykorzystano 10 osób, które fizycznie wykonywały gesty w obszarze roboczym elektrod. W rezultacie potwierdzono przydatność zaproponowanej metody akwizycji gestów. W poniższej tabeli przedstawiono wyniki skuteczności uzyskane dla rozpoznawanych gestów.

Tabela 2. Skuteczność rozpoznawania gestów wykonywanych przez rzeczywistych użytkowników – dane do uczenia sieci neuronowej zebrano wykorzystując wyłącznie manipulator.[2]

	Swipe left	Swipe right	Swipe up	Swipe Down	Tap	sumarycznie
Dokładność	96 %	91 %	60 %	76 %	78 %	80.2 %

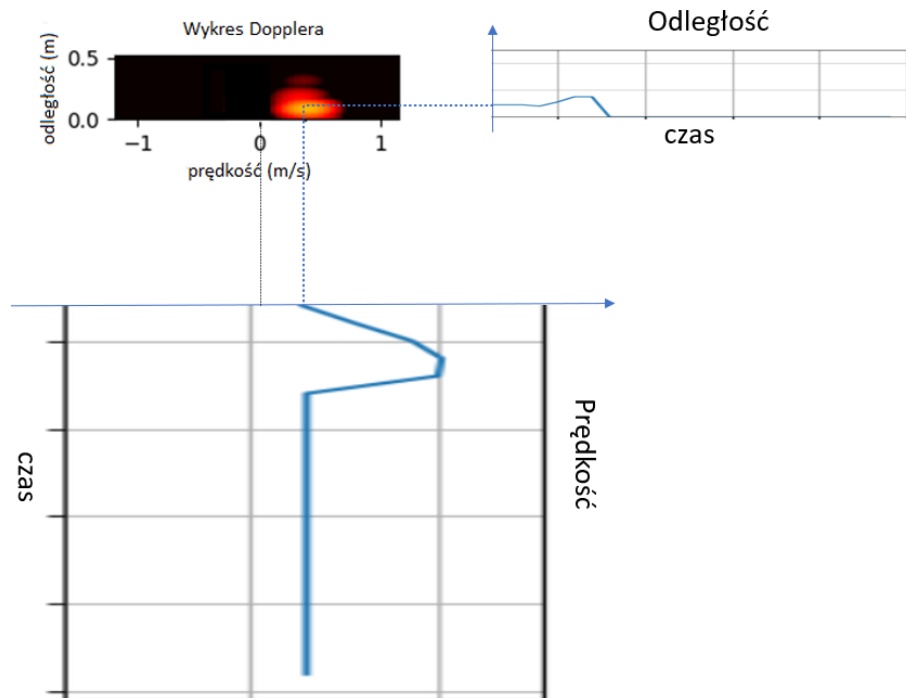
W tabeli nr 3 przedstawiono porównanie metod zbierania gestów podczas prac z systemem pojemnościowym. Przedstawione wnioski potwierdzają przydatność zaproponowanej metody zbierania gestów. Może być ona wykorzystana jako wyłączone źródło zestawów danych uczących a także w połączeniu z danymi zbieranymi metoda tradycyjną.

Tabela 3. Porównanie metod akwizycji danych – automatycznej i manualnej [2]

Kategoria	Metoda manualna	Metoda Automatyczna
Czas zbierania danych	Czasochłonna	Krótszy, możliwość pracy bez nadzoru
Zaangażowanie osoby nadzorującej	Podczas całego procesu	Podczas przygotowania oraz startu
Spójność zestawu danych z aplikacją docelową	Bardzo wysoka	Wysoka
Nakład pracy związany z przygotowaniem środowiska testowego	Średni	Wysoki
Liczba osób wykonujących gesty	Wysoka	Brak
Różnorodność danych	Zależy od użytkowników	Zależy od zastosowanych algorytmów randomizacji
Liczba gestów podczas przeprowadzonych testów	123 gesty/h	427 gestów/h

Kolejnym zagadnieniem, które jest efektem prac nad systemem radarowym, jest autorski sposób przetwarzania danych zebranych bezpośrednio z radaru. Zaproponowane podejście wykorzystuje znacznie mniejszą liczbę danych wejściowych do modelu sieci neuronowej, przez co zwiększa możliwości uruchomienia niniejszego rozwiązania na systemie wbudowanym nie posiadającym ogromnych zasobów obliczeniowych oraz pamięci. Główną różnicą w stosunku do innych tego typu rozwiązań [3][4], jest wykorzystanie tylko wartości maksymalnej uzyskiwanej z mapy dopplera, zamiast pełnej mapy lub całej kolumny zawierającej wartość maksymalną. Pomimo wykorzystania kilkunastokrotnie mniejszej ilości danych, uzyskana skuteczność rozpoznawania gestów była tylko o 10% procent niższa. Jednak w

przypadku wykorzystania danych z radaru oraz czujnika pojemnościowego skuteczność rozpoznawania gestów była niemal identyczna, podczas gdy nadal wykorzystywano kilkunastokrotnie mniejszą liczbę danych. Na poniższym rysunku przedstawiono sposób ekstrakcji danych wykorzystywanych do uczenia z wykresu Dopplera.



Rysunek 3. Ekstrakcja odległości oraz prędkości z Mapy Dopplera

Przeprowadzone badania i eksperymenty potwierdziły postawione tezy. Dodatkowo zaproponowano nowe rozwiązania w zagadnieniach związanych z tematyką systemów rozpoznających gesty, a także uczenia maszynowego. Przedstawione wyniki analiz mogą zostać wykorzystane przy opracowywaniu sytemów wykorzystywanych do podobnych celów.

6. Literatura

- [1] H. Ismail Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P. A. Muller i F. Petitjean, „InceptionTime: Finding AlexNet for time series classification” *Data Mining and Knowledge Discovery*, tom 34, nr 6, pp. 1936-1962, 11 2020.
- [2] P. Reczek, J. Panczyk, A. Wetula i Młyniec Andrzej, „Data Collection Automation in Machine Learning Process Using Robotic Manipulator” w *IFIP Advances in Information and Communication Technology*, Leon, 2023.
- [3] M. Chmurski, G. Mauro, A. Santra, M. Zubert i G. Dagasan, „Highly-optimized radar-based gesture recognition system with depthwise expansion module” *Sensors*, tom 21, nr 21, 11 2021.
- [4] G. Mauro, M. Chmurski, L. Servadei, M. Pegalajar-Cuellar i D. P. Morales-Santos, „Few-Shot User-Definable Radar-Based Hand Gesture Recognition at the Edge” *IEEE Access*, tom 10, pp. 29741-29759, 2022.