Warszawa, dn. 17 listopada 2024

dr hab. inż. Marcin Iwanowski, prof. uczelni
Politechnika Warszawska, Wydział Elektryczny
Instytut Sterowania i Elektroniki Przemysłowej

Recenzja rozprawy doktorskiej mgr inż. Katarzyny Gościewskiej pt. „Metoda rozpoznawania akcji wykorzystująca analizę kształtu na potrzeby wizyjnych systemów wspomagania opieki nad osobami starszymi"

1 Zawartość rozprawy

Recenzowana rozprawa liczy 119 stron, w tym 99 stron tekstu zasadniczego. Została napisana w języku polskim, składa się z nienumerowanego wprowadzenia, czterech rozdziałów, nienumerowanych zakończenia, spisów rysunków i tabel oraz bibliografii.

Wprowadzenie rozpoczyna się od naszkicowania problemu starzejącego się społeczeństwa i wynikających z tego rosnących potrzeb w zakresie opieki nad osobami starszymi. Następnie omwione są zagadnienia techniczne stojące u podstaw systemów wspomagania opieki ze szczególnym uwzględnieniem systemów wizyjnych – pozyskiwania obrazów i ich przetwarzania. Przedstawiona jest także motywacja Autorki do zajęcia się tytułowym tematem oraz układ rozprawy. Na koniec sformułowana jest teza rozprawy, która brzmi: „Wykorzystanie cech kształtu sylwetek wyekstrahowanych z sekwencji wideo pozwoli na opracowanie skutecznej metody rozpoznawania akcji wykonywanych przez pierwszoplanowe postacie ludzkie na potrzeby zastosowania w systemach wspomagania opieki bazujących na monitoringu wizyjnym" oraz określone na potrzeby jej wykazania szczegółowe cele pracy.

Rozdział pierwszy zawiera ogólne omówienie zagadnień wspomagania opieki nad osobami starszymi w kontekście wykorzystywanych rozwiązań technicznych. W pierwszym punkcie omówione zostaje społeczne i demograficzne tło problemu badawczego, ze zaakcentowaniem jego znaczenia we współczesnym świecie. Następnie Autorka przechodzi do przedstawienia systemów technicznych wspierających opiekę nad osobami starszymi. W punkcie trzecim opisane są systemy tego typu wykorzystujące widzenie komputerowe, z uwzględnieniem stosowanych w nich algorytmów.

Rozdział drugi jest poświęcony metodom przetwarzania i analizy obrazu na potrzeby identyfikacji akcji w sekwencjach wideo. Na początku przedstawione są podstawowe definicje i taksonomie. W punkcie drugim Autorka omawia metody nazwane tradycyjnymi, choć lepszą nazwą byłaby „metody klasyczne", w których pierwszy etap projektowania algorytmu detekcji dla konkretnych rodzajów obrazów i obiektów, polega na analizie i świadomym wyborze cech. W trzecim punkcie są omówione metody aktualnie dominujące, w których selekcja cech następuje w sposób automatyczny – w pracy metody te są nazwane metodami opartymi na uczeniu się cech.

Rozdział trzeci jest jednym z dwóch zasadniczych w opiniowanej rozprawie i zawiera omówienie proponowanej metody. W pierwszym punkcie omówione są dane wejściowe i sposób ich przetwarzania wstępnego, w tym sposoby określania obszaru, w którym znajduje się poruszająca się sylwetka. W drugim punkcie jest przedstawiony schemat proponowanej metody składającej się z kilku etapów. W punkcie trzecim przedstawione są podstawowe algorytmy ekstrakcji cech i tworzenia deskryptorów kształtu, miary dopasowania, reprezentacji sekwencji czasowej deskryptorów kształtu oraz klasyfikacji.

Drugim z zasadniczych rozdziałów pracy jest rozdział czwarty, w którym omówiono wykonane eksperymenty oraz ich wyniki. W pierwszym punkcie przedstawiono podstawowe założenia i procedurę testową. To ważny punkt pracy, w którym znajdują się opisy dwóch ogólnodostępnych zbiorów danych: Weizmann oraz AMASS wraz z procedurami przetwarzania wstępnego oraz ogólnego planu eksperymentów przeprowadzanych na każdym ze zbiorów. W kolejnych punktach omówiono poszczególne eksperymenty wykorzystujące bazę Weizmann oraz: złożone deskryptory kształtu i standardowy klasyfikator, proste deskryptory kształtu obliczane na podstawie sylwetki człowieka, wpływ uogólnienia cech kształtu na skuteczność metody, klasyfikator w postaci prostej sieci neuronowej. W punkcie siódmym zawarto porównanie zaproponowanego podejścia z metodami znanymi z literatury. Punkt ósmy zawiera opis eksperymentu z zaproponowaną metodą zastosowaną do klasyfikacji akcji w sekwencjach ze zbioru AMASS.

W zakończeniu, Autorka podsumowuje przeprowadzone badania.

Na końcu pracy zamieszczono spisy rysunków i tabel oraz obszerną bibliografię zawierającą 218 pozycji literatury.

2 Ocena merytoryczna pracy

Praca jest poświęcona zagadnieniu rozpoznawania rodzaju czynności wykonywanej przez człowieka na podstawie sekwencji wideo pokazującej daną czynność. Sekwencje akcji będące przedmiotem badań składają się z masek binarnych ukazujących sylwetkę osoby poruszającej się i powstały z oryginalnych sekwencji obrazów barwnych w wyniku przetwarzania wstępnego. Sekwencje oryginalne zawierają poruszającego się na różne sposoby człowieka znajdującego się na jednolitym tle. W pracy wykorzystano dwie ogólnodostępne bazy sekwencji: Weizmann, zawierającą rzeczywiste nagrania oraz bazy AMASS, w której skład wchodzą sekwencje wygenerowane sztucznie z użyciem grafiki komputerowej. W każdej z tych baz znajdują się sekwencje pokazujące osoby wykonujące określone czynności, nazywane w pracy akcjami. Celem pracy było zbadanie możliwości detekcji akcji na podstawie sekwencji z użyciem klasycznego podejścia opartego na selekcji cech. W dalszym planie przeprowadzonych badań znajduje się zastosowanie proponowanej metodologii w systemach wizyjnych wykorzystywanych do opieki nad osobami starszymi.

Zaproponowane podejście dotyczy klasycznych metod przetwarzania obrazów. Podejście to jest aktualnie rzadziej stosowane, gdyż ustąpiło ono miejsca metodom automatycznej selekcji (uczenia się) cech opartym na uczeniu głębokim. Jednak metody klasyczne są szybsze i w wielu przypadkach zupełnie wystarczające do realizacji zadań detekcji. Podjęcie się zatem badań nad tą grupą metod uważam za właściwe.

W ramach eksperymentów zbadano różne kombinacje cech z użyciem dwóch klasyfikatorów: najbliższego sąsiada i prostej sieci neuronowej. Wybrane wyniki eksperymentów zostały porównane z wynikami raportowanymi w literaturze.

Autorka zaproponowała w pracy (punkt 3.2) schemat działania w ramach którego realizowane są kolejno następujące kroki:

1. przetwarzanie wstępne klatki sekwencji w celu uzyskania obrazu binarnego – maski sylwetki,
2. wyznaczenie środka ciężkości maski i deskryptora kształtu (dobór cech wykorzystanych w deskryptorze jest przedmiotem badań),
3. łączenie deskryptorów kształtów wszystkich klatek sekwencji w jeden wektor cech i jego normalizacja,
4. przekształcenie wektora cech,
5. klasyfikacja.

Badania nad doborem cech przeprowadzono zgodnie z ustaloną procedurą (przedstawioną w punkcie 4.1) w czterech wątkach. W ramach pierwszego (punkt 4.2) zbadano złożone deskryptory kształtu. W drugim wątku (punkt 4.3) badaniu podlegały proste deskryptory kształtu oryginalnego, zaś w trzecim (punkt 4.4) te same deskryptory dla otoczki wypukłej i prostokąta ograniczającego. Badanie w tych trzech wątkach były przeprowadzone dla bazy Weizmann oraz klasyfikatora najbliższego sąsiada. Badania z pierwszego i drugiego wątku zostały zrealizowane także w wariancie z klasyfikatorem w postaci prostej sieci neuronowej (co opisano w punkcie 4.5). Zaproponowane podejście zostało skonfrontowane z innymi metodami (punkt 4.6). Z porównania tego wynika, że zaproponowane podejście przewyższa cześć spośród wymienionych metod znanych z literatury. Istnieje także, wymieniona w pracy grupa metod, dla których wyniki uzyskane przez Autorkę są porównywalne. Jednak, co słusznie zostało zaakcentowane w pracy, podejście proponowane jest prostsze, co przy porównywalnych wynikach stanowi o jego przewadze. W osobnym wątku badań zweryfikowano podejście wykorzystujące proste deskryptory kształtu i klasyfikator w postaci prostej sieci neuronowej uczonego na bazie AMASS (punkt 4.7).

W powyższych wątkach przeprowadzono znaczącą liczbę eksperymentów w których wykorzystywano różne cechy do otrzymania deskryptora sylwetki, sposoby tworzenia wektorów cech oraz ich porównywania (dopasowania). Zaproponowane podejście pozwoliło na określenie optymalnych kombinacji cech do realizacji zadania klasyfikacji akcji.

Zaproponowany schemat jest właściwy, zaś autorski pomysł na łączenie deskryptorów w pojedynczy wektor cech i jego przekształcenie transformacją Fouriera lub jej wariantem – periodogramem w celu uzyskania ostatecznego wektora cech sekwencji, jest ciekawy i – w zaprezentowanej formie – wraz z testami różnych cech kształtu maski sylwetki może być uznany za osiągnięcie Autorki. Przeprowadzone eksperymenty skutecznie potwierdziły przydatność zaproponowanego podejścia do klasyfikacji akcji wyraźnie widocznych sylwetek w ruchu w testowych sekwencjach obrazów. Wyniki dowodzą przydatności proponowanego podejścia do wspomagania opieki nad osobami starszymi.

Prace badawcze opisane w pracy zostały przeprowadzone starannie i zgodnie z właściwie przygotowanym planem badawczym. Praca jest także bardzo staranna pod względem edytorskim, z dobrze dobranym materiałem ilustracyjnym. Zawiera ponadto obszerny przegląd literatury z przedmiotowego zakresu. Na uwagę zasługuje rozdział 3.3, w którym znajduje się omówienie klasycznych deskryptorów kształtu. Stanowi ono dobry materiał referencyjny.

3 Uwagi krytyczne

Lektura pracy pozwoliła na sformułowanie następujących uwag dyskusyjnych:

1. **Wizualizacja i analiza eksploracyjna przestrzeni cech.** Zaproponowane podejście polega na wykorzystaniu klasycznych cech kształtu. Cechy te są albo skalarami, albo tworzą wektory o stosunkowo niewielkich rozmiarach. Poszczególne klatki sekwencji ruchu są zatem przekształcane do reprezentacji wektorowej i w efekcie stają się punktami w wielowymiarowej przestrzeni cech. Dobrym rozwiązaniem w takich przypadkach jest przeprowadzenie eksploracyjnej analizy danych pozwalającej na określenie zmienności wartości poszczególnych cech oraz ich wizualizacji. W szczególności przydatne byłoby przeprowadzenie takich analiz dla poszczególnych klas (akcji) oraz skonfrontowanie ich ze sobą. Proces taki przeprowadzony zarówno dla deskryptorów sylwetki jak i dla wektora cech sekwencji, pozwoliłby na uzupełniającą analizę separowalności klas (akcji) w zależności od rodzaju deskryptora sylwetki i metody pozyskiwania wektora cech sekwencji. Dopełniałaby ona zawartą w pracy analizę bazująca na wynikach klasyfikacji. Jakie narzędzia mogłyby zostać w tym celu wykorzystane ?

2. **Dobór klasyfikatorów.** W pracy wykorzystano dwa klasyfikatory – najbliższego sąsiada i prostą sieć neuronową. Każdą z nich wykorzystano przy tym w innym eksperymencie, dla innego sposobu pozyskiwania wektorów cech. Dobrym uzupełnieniem eksperymentów byłoby wykorzystanie obu we wszystkich przeprowadzonych eksperymentach, co pozwoliłoby na ich porównanie. Poza tym dobór tych klasyfikatorów jest dość arbitralny i niedostatecznie szczegółowo uzasadniony. Jakie inne klasyfikatory mogłyby być użyte i jak mogłoby to wpłynąć na ostateczne wyniki ?

3. **Parametryzacja metod przetwarzania wstępnego.** Zaproponowany schemat proponowanej metody zakłada, i słusznie, konieczność wstępnego przetwarzania obrazów – klatek sekwencji. Kolejne elementy przetwarzania wstępnego przedstawiono w rozdziale 4 na str. 71 dla zbioru Weizmann, oraz na str. 75 dla zbioru AMASS. Wśród tych elementów są operacje parametryczne, których parametry (próg binaryzacji, rozmiary elementów strukturujących w operacjach morfologicznych) zostały dobrane empirycznie tak by, cyt. "sylwetka nie straciła szczegółów kształtu". Intuicyjnie jest to dobre podejście. Jednak takie rozwiązanie jest w zasadniczym stopniu zależne od cech charakterystycznych konkretnych sekwencji wykorzystywanych w eksperymentach. Zresztą potwierdza to niejako fakt, iż inna procedura przetwarzania wstępnego została zaproponowana dla bazy Weizmann, inna zaś dla bazy AMASS. Od algorytmów wizyjnych, szczególnie przeznaczonych do praktycznych zastosowań, należałoby oczekiwać jak największych zdolności generalizacyjnych, nawet kosztem zbyt dużego dopasowania do zbioru uczącego. W jaki zatem sposób można by zapewnić w proponowanym podejściu zwiększenie zdolności do generalizacji metody w zakresie przetwarzania wstępnego ?

4. **Ograniczenie jedynie do jednego rodzaju aktywności osób starszych.** Tytuł pracy oraz rozdział 1 wprost odnoszą się do systemów, w szczególności wizyjnych, przeznaczonych do wspomagania opieki nad osobami starszymi. W przypadku takich osób można wyróżnić dwa aspekty aktywności fizycznej: pierwszy to aktywność w codziennym funkcjonowaniu, zaś druga – to ćwiczenia fizyczne. W pierwszych przypadku od systemu wizyjnego ułatwiającego opiekę należałoby oczekiwać detekcji specyficznych zdarzeń ruchowych takich jak np. upadek,

czy określony sposób wykonywania kroków wskazujący na zaburzenia równowagi itp. W pracy został uwzględniony jedynie drugi rodzaj aktywności – co ilustrują akcje wykonywane przez osoby uwidocznione w sekwencjach wideo – z pominięciem pierwszego wspomnianego aspektu. Wydaje się, że dobrym rozszerzeniem pracy byłyby testy metody na aktywnościach typowych dla codziennego funkcjonowania osób starszych. Dlaczego ten rodzaj aktywności nie został w pracy uwzględniony ?

5. **Nieobecność metod uczenia głębokiego.** Zauważalny w pracy jest brak eksperymentów z najnowszymi metodami uczenia głębokiego. Szczególnie w ujęciu takim jak zaproponowała to Autorka tj. detekcji akcji na sekwencjach nagranych w warunkach laboratoryjnych z dobrze widoczną sylwetką osoby wykonującej ruchy. Metody uczenia głębokiego, o których zresztą sama Autorka wspomina w rozdziale 2.3, są doskonałym narzędziem pozwalającym na automatyczny dobór cech – są przykładem metod opartych na uczeniu się cech. Szczególnie ciekawe byłoby w tym kontekście zastosowanie sieci splotowych na podobrazie zawierającym wyciętą sylwetkę poruszającego się człowieka. Otrzymalibyśmy wówczas wektory cech, których sekwencje, podobnie jak w zaproponowanym podejściu, można by wykorzystać do pozyskania reprezentacji akcji poddawane następnie klasyfikacji zgodnie z zaproponowanym schematem. Z kolei inne sieci uczenia głębokiego – sieci rekurencyjne typu np. LSTM lub GRU można by wykorzystać jako alternatywny do zaproponowanego (klasyfikacja reprezentacji akcji poprzedzonej łączeniem deskryptorów) sposób przetwarzania sekwencji wektorów cech w celu detekcji rodzaju akcji. Tego typu eksperymenty pozwoliłyby na konfrontację metod klasycznych, których praca faktycznie dotyczy, z metodami uczenia głębokiego. W jakich przypadkach metody uczenia mogłyby poprawić efektywność detekcji akcji ?

4 Uwagi redakcyjne

Z uwagi na dużą staranność edytorską, liczba znalezionych niedociągnięć jest niewielka. W trakcie ~~lektury pracy zwróciły moją uwagę następujące usterki~~:

- W wielu miejscach pracy Autorka powołuje się na pozycje literatury zawierające faktycznie omówienia danych metod czy algorytmów. W ich miejsce należałoby raczej podać tam odniesienia do oryginalnych źródeł.

- Odniesienia do elementów numerowanych tekstu – rysunków, tabel itp, powinny zawierać odpowiednie słowo („rysunek", „rys.", itp.) pisane z małej, a nie wielkiej litery. O ile oczywiście nie jest to pierwsze słowo w zdaniu.

- Na stronie 35 znajduje się lista przedstawiająca modele głębokiego uczenia (górna część strony) – głębokie sieci neuronowe zostały w niej przypisane do grupy modeli nadzorowanych. Nie jest to właściwe, gdyż modele takie są wykorzystywane także w uczeniu nienadzorowanym np. jako kodery (enkodery). Słowo „głęboki" w tej nazwie odnosi się do struktury sieci, a nie jej przeznaczenia.

- Podpisy pod niektórymi rysunkami i tabelami są za długie (np. rys. 3.5, 3.7, tab. 4.10). Lepszym rozwiązaniem jest pozostawienie krótkiego podpisu z przeniesieniem szczegółowego opisu treści do tekstu zasadniczego.

- W kilku miejscach pracy (np. str. 79, pierwszy akapit punktu 4.2) Autorka używa czasownika „dopasowywać" w odniesieniu do określania stopnia podobieństwa pomiędzy dwoma deskryptorami. O ile w kontekście miary, określenie używające rzeczowników „miara dopasowania", jest zasadne, o tyle jako czynność raczej odnosi się ono do dokonywania zmian, w efekcie których jeden wektor staje się bardziej podobny do drugiego. Tymczasem sens użycia ~~tego słowa w pracy odnosi się do czynności biernej analizy dwóch wektorów~~ (tj. ~~bez zmiany~~ zawartości żadnego z nich) w celu określenia podobieństwa. Jest to słuszne podejście, jednak lepszym w tym przypadku określeniem byłoby „porównywanie" deskryptorów.

5 Podsumowanie

Podsumowując, niezależnie od zawartych w niniejszej opinii uwag krytycznych, które w większości są elementem dyskusji nad pracą oraz sugestią ewentualnych dalszych prac badawczych, uważam, że Autorka pracy wykazała kompetencje w wybranym przez siebie obszarze badawczym. Recenzowana rozprawa zawiera oryginalne rozwiązania problemów naukowych oraz pokazuje umiejętności Autorki w zakresie samodzielnego prowadzenia pracy naukowej. **Całościowa ocena rozprawy jest pozytywna**. Uważam, że rozprawa spełnia wymagania stawiane rozprawom doktorskim przez stosowne, aktualnie obowiązujące, akty prawne. Dlatego **wnioskuję o dopuszczenie rozprawy doktorskiej mgr inż. Katarzyny Gościewskiej do publicznej obrony**.