



Prof. dr hab. Ewa Grabska
Zakład Projektowania i Grafiki Komputerowej
Wydział Fizyki, Astronomii i Informatyki Stosowanej
Uniwersytet Jagielloński

Kraków, 30. 09. 2024

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: Klasyfikacja zachowań postaci ludzkiej z wykorzystaniem uczenia maszynowego na podstawie trajektorii punktów charakterystycznych w reprezentacji 3D i 2D.

Autor rozprawy: mgr inż. Magdalena Pawlyta

Promotor rozprawy: prof. dr hab. inż. Konrad Wojciechowski

Promotor pomocniczy rozprawy: dr inż. Przemysław Skurowski

Informacje ogólne

Oceniana rozprawa doktorska mgr inż. Magdaleny Pawlyty zatytułowana „Klasyfikacja zachowań postaci ludzkiej z wykorzystaniem uczenia maszynowego na podstawie trajektorii punktów charakterystycznych w reprezentacji 3D i 2D” stanowi opracowanie o objętości 176 stron, zawierające 163 rysunki, 34 tabele oraz spis literatury obejmujący 98 pozycji.

I. O przedmiocie rozprawy

Współcześnie, system wizyjny stanowi nieodzowne narzędzie dostarczające danych zastępujących i/lub wspomagających percepcję wizualną. Tematyka rozprawy dotyczy jakości rozpoznawania zachowania postaci ludzkiej z użyciem systemów wizyjnych i wykorzystaniem technologii przechwytywania ruchu (ang. motion capture).

Zachowanie postaci jest charakteryzowane przez sekwencję akcji związanych z ruchem. Klasyfikacja odbywa się na podstawie danych specyfikujących wykonywane akcje. W rozprawie zastosowano system śledzenia ruchu, który do wyznaczania akcji postaci używa rozmieszczonych na postaci znaczników zwanych markerami. Taki system markerowy jest wyposażony w odpowiednią, w zależności od zadania liczbę kamer stereowizyjnych emitujących światło podczerwone. Kamery rejestrują odbite światło. Obrazy z kamer służą do wyznaczenia trajektorii wybranych punktów sylwetki w przestrzeni trójwymiarowej. Trajektorie punktów są źródłem danych dla klasyfikatorów. Do klasyfikacji wykorzystuje się narzędzia uczenia maszynowego.

Tematyka badań naukowych, którymi zajmuje się Doktorantka jest ważna z punktu widzenia informatyki, między innymi dotyczy zagadnień rozpoznawania obrazów. Znajduje również zastosowania w wielu innych dziedzinach. Monitoring wizyjny jest obecnie jednym z najszerzej wykorzystywanych środków poprawiających bezpieczeństwo i porządek publiczny. Od wielu lat w medycynie analiza ruchu pacjentów z użyciem systemów wizyjnych jest jednym z elementów oceny stanu zdrowia. Znane są również zastosowania w sporcie oraz opiece domowej.

II. O postawionych w rozprawie celach

Cel jaki stawia sobie Doktorantka to badanie zależności jakości rozpoznawania zachowania postaci ludzkiej od systemu kamer i ich parametrów zarówno wewnętrznych, jak i zewnętrznych. Dane do klasyfikacji zachowania są uzyskane z wykorzystaniem metody przechwytywania ruchu.

Pierwszym etapem do przeprowadzonych eksperymentów związanych z zachowaniem postaci był wybór odpowiednich akcji. Autorka zmniejszyła liczbę akcji zwykle używanych w metodach rozpoznawania zachowania postaci usuwając akcje nie mające istotnego znaczenia dla rozważanych charakterystyk ruchu.

Następny etap polegał na wyborze metody uzyskiwania oraz reprezentacji danych. W tym celu wykorzystano technologię przechwytywania ruchu (motion capture). Na poziomie modelu, z trajektorii wybranych punktów sylwetki w przestrzeni trójwymiarowej odpowiadających systemowi stereowizyjnemu, w kolejnym etapie za pomocą rzutów perspektywicznych utworzona została baza danych ruchu w przestrzeni dwuwymiarowej. Wymagało to opracowania kamery wirtualnej ustawianej w różnych lokalizacjach na scenie wirtualnej.

Do eksperymentów na etapie klasyfikacji wykorzystano dwie architektury głębokiego uczenia sieci neuronowych CNN oraz LSTM. Doktorantka rozważała wyniki klasyfikacji w zależności od wymiaru przestrzeni.

III. Przegląd zawartości rozprawy

W **Rozdziale 1** Autorka dokonała przeglądu ogółu problemów dotyczących analizy i klasyfikacji zachowania człowieka obserwowanego z pojedynczej kamery wizyjnej. Następnie przedstawiła zakres wybranej tematyki, jako przedmiotu swoich badań oraz sformułowała tezę rozprawy doktorskiej.

Rozdział 2 prezentuje różne aspekty zachowania postaci ludzkiej, które są zależne od stopnia złożoności wykonywanych akcji. Podjęcie się problemu klasyfikacji zachowania wymaga rejestracji danych dotyczących wykonywanych akcji oraz ekstrakcji cech. Ważną rolę odgrywa odpowiednia selekcja cech oraz ich reprezentacja. Doktorantka podkreśla znaczenie wykorzystania metod głębokiego uczenia do ekstrakcji cech.

W dalszej części rozdziału Autorka dokonała wnikliwego przeglądu publicznych baz danych ruchu. Różne bazy są zorientowane na rozpoznawanie różnych rodzajów akcji. Klasyfikacja akcji zależy od parametrów kamery oraz od orientacji danej osoby względem kamery. Ze względu na tematykę rozprawy proponowano zbadanie wpływu rzutowania perspektywicznego na sam rodzaj akcji oraz sposób jej wykonania za pomocą nowo utworzonej bazy danych z użyciem technologii przechwytywania ruchu.

W **Rozdziale 3** Autorka opisała rodzaj automatycznego uczenia się za pomocą głębokiej sieci neuronowej (ang. Deep Neural Network – DNN)). Następnie scharakteryzowała dwa warianty sieci LSTM (ang. Long Short-Term Memory). Zaprezentowane sieci zostały wybrane jako klasyfikatory do zbadania aspektów wpływających na jakość wybranych akcji.

Na początku **Rozdziału 4** zaprezentowana została metoda przechwytywania obrazu, znana głównie jako narzędzie do tworzenia animacji komputerowych, ale również stosowana w medycynie czy sporcie do rozpoznawania i klasyfikacji akcji postaci. Dane wykorzystane w systemie akwizycji ruchu w niniejszej rozprawie doktorskiej pochodzą z laboratorium Human Motion Laboratory (HML) mieszczącego się w Centrum Badawczo Rozwojowym Polsko Japońskiej Akademii Technik Komputerowych (CBR-PJATK) w Bytomiu. W dalszej części rozdziału dokonano wyboru odpowiednich akcji prostych. Po dogłębnej analizie Autorka zmniejszyła ich liczbę w stosunku do liczb akcji zwykle używanych w metodach rozpoznawania zachowania postaci. W części końcowej rozdziału opisała metody standaryzacji danych.

W **Rozdziale 5** Doktorantka przedstawiła klasyfikację zachowania postaci w przestrzeni trójwymiarowej za pomocą głębokich sieci neuronowych DNN oraz LSTM. Ze względu na dużą liczbę markerów liczba danych wejściowych została zredukowana w sposób pozwalający zachować dokładność sieci. Na zakończenie rozdziału dokonano porównania wyników dla obu typów sieci.

W **Rozdziale 6** Autorka przeprowadziła eksperymenty obejmujące klasyfikację wybranych akcji w przestrzeni dwuwymiarowej. Dodatkowo, zbadała skutecz-

ność rzutowania perspektywicznego na ogólną dokładność klasyfikacji oraz na rozpoznawalność poszczególnych akcji.

Rozdział 7 zawiera podsumowanie oraz pokazuje wybrane kierunki kontynuacji badań.

IV. Ocena wyników rozprawy

Podjęcie się ambitnego zadania, polegającego na analizie zależności jakości rozpoznawania akcji postaci od systemu kamer i ich parametrów zewnętrznych i wewnętrznych zasługuje na uznanie. Doktorantka wykazała, że zna poruszany temat oraz ma własne oryginalne pomysły, które potrafi zrealizować. W sposób właściwy została opracowana baza trójwymiarowych danych akcji ruchu z algorytmem podziału bazy danych na dowolną liczbę podzbiorów typów ruchu umożliwiającą operację krzyżowania i walidację. Należy podkreślić, że dokonano doboru akcji postaci stanowiących podstawę eksperymentu po dogłębnej analizie, jak również opisy wybranych akcji zostały zaprezentowane z niezwykłą starannością. Na poziomie klasyfikacji Autorka trafnie wybrała architektury sieci neuronowych głębokiego uczenia do zbadania aspektów wpływających na jakość wybranych akcji. Powiększenie zakresu badań o jednokamerowy system nagrań z trajektoriami dwuwymiarowymi wraz z klasyfikacją za pomocą sieci neuronowych głębokiego uczenia jest nowatorskie i stanowi podstawę ciekawego problemu badawczego, na przykład na temat ułatwienia analizy ruchu postaci.

Układ rozprawy jest prawidłowy.

Cel rozprawy został osiągnięty.

Rozprawa ma również i słabsze punkty.

Uwagi szczegółowe:

1.

Brak wyjaśnień niektórych symboli we wzorach,
Przykłady: wzór 3 str. 18, wzór 8 str.19.

2.

Strona 25:

Jest:

ilość kamer, ilość znaczników

Powinno być:

liczba kamer, liczba znaczników (rzeczowniki *kamera* i *znacznik* są przeliczalne)

3.

Strona 32:

Wzór 17 przedstawia przykład transformacji obrotu punktów w przestrzeni trójwymiarowej w prawidłowej reprezentacji z użyciem współrzędnych jednorodnych. Dlaczego użyto zapisu $1'$ po lewej stronie wzoru?

Błędne użycie 1' zamiast stałej o wartości 1 zawierają wszystkie wzory transformacji w przestrzeni trójwymiarowej za pomocą współrzędnych jednorodnych.

4.

Strona 47:

Zapis wzoru PBA na poziom błędu akcji (35) jest niepoprawny. Składowe wzoru PBO_i nie są zdefiniowane, dodatkowo pomyłona jest nazwa składowych (POB).

Zapis $PBO = \frac{\text{Liczba błędnych klasyfikacji osoby}}{\text{Wszystkie akcje osoby}}$ nie jest wzorem matematycznym.

Ten typ błędu polegający na braku formalnego zdefiniowania składowych występujących we wzorach można znaleźć również w innych prezentowanych wzorach w rozprawie.

Konkluzja

W sumie nie mam jednak poważniejszych zastrzeżeń do rozprawy i oceniam ją zdecydowanie pozytywnie. Oceniana rozprawa doktorska spełnia wymagania, jakie ustawa o stopniach i o tytule naukowym przewiduje dla rozpraw doktorskich w dziedzinie nauk inżyniersko-technicznych w dyscyplinie: Informatyka Techniczna i Telekomunikacja oraz na tej podstawie wnoszę o dopuszczenia jej Autorki – mgr inż. Magdaleny Pawłyty – do publicznej obrony.

