



Politechnika Łódzka

Instytut Informatyki

Łódź, 2 września 2022

dr hab. inż. Adam Wojciechowski, prof. PŁ
Instytut Informatyki
Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej
Politechnika Łódzka

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: **Learning Scene Dependent Models of Human Actions Behavior from Surveillance Video**

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

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

Promotor dodatkowy: **dr inż. Jakub Segen**

Niniejsza recenzja została sporządzona na wniosek Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja, Polsko-Japońskiej Akademii Technik Komputerowych, która powołała mnie na recenzenta w przewodzie doktorskim Pana mgr inż. Marka Kulbackiego, wszczętym w dyscyplinie Informatyka Techniczna i Telekomunikacja.

1. Zakres tematyczny rozprawy

Tematyka rozprawy dotyczy problematyki śledzenia i rozpoznawania zachowań postaci w czasie rzeczywistym bazując na analizie wizualnych cech sekwencji obrazów pochodzących z systemów monitoringu przestrzeni miejskich. Szczególną uwagę przywiązano do efektywności, wydajności i przemysłowej przydatności opracowanych metod, których zadaniem jest śledzenie obrazów z kamer w warunkach rzeczywistych, czego wymagają profesjonalne systemy monitoringu. Chociaż zagadnienia wizji komputerowej poświęcone śledzeniu i wykrywaniu zachowań postaci są przedmiotem badań od wielu lat to efektywne, skuteczne i realizowalne w trudnych warunkach środowiskowych, wielokamerowe monitorowanie zachowań postaci wciąż stanowi wyzwanie dla obecnych systemów komputerowych. Warto podkreślić, że nominalnie najwyższe skuteczności publikowanych w literaturze metod dotyczą głównie kontrolowanych warunków laboratoryjnych lub metod, których trening był dobrze zdefiniowany i wsparty istotną mocą obliczeniową. Warunki rzeczywiste natomiast, w których mamy do czynienia z niesprzyjającym oświetleniem sceny, dużą liczbą ruchomych obiektów o nieznanym kształcie, lub nawet częściowo przysłoniętych, stanowią wciąż duże wyzwanie dla algorytmów rozpoznawania i klasyfikacji ruchów człowieka. Systemy, które dodatkowo mają działać w czasie rzeczywistym, obarczone są zatem dodatkowymi ograniczeniami wydajnościowymi i wciąż stanowią wyzwanie dla współczesnej nauki.

W kontekście powyższych uwag, zarówno zakres jak i problematykę rozprawy doktorskiej należy uznać za ciekawą i aktualną, zarówno pod względem teoretycznym jak również praktycznym.

2. Zawartość i układ pracy

Rozprawa doktorska obejmuje 336 stron i składa się z 12 rozdziałów podzielonych na 3 części oraz bibliografii zawierającej 314 pozycji, jak również dodatku ze schematami, najważniejszymi fragmentami kodów programów komputerowych oraz instrukcją opisującą kluczowe komponenty

systemu SAVA – systemu monitorowania miejskiego tworzono w ramach projektu NCBR, którego Doktorant jest współtwórcą.

Za rozdział wprowadzający w problematykę badań, należy uznać rozdział numer 4. Przedstawia on wykorzystywaną w pracy terminologię, w tym kluczową z punktu widzenia pracy taksonomię aktywności człowieka według Aggarwala (akcja, interakcja, gest, grupa aktywności), z czego Autor rozprawy skoncentrował się na „zachowaniach/akcjach” człowieka. Motywacją do podjęcia badań jest stworzenie alternatywnej dla istniejących w literaturze reprezentacji zachowań/akcji człowieka, która łączyłaby w sobie cechy metod opartych o globalne charakterystyki ruchu z własnościami metod opartych o model strukturalnej/hierarchicznej reprezentacji ciała ludzkiego. Zaproponowane rozwiązanie ma wykorzystywać nową, strukturalną reprezentację ruchu, ale bez konieczności odwoływania się do modelu szkieletowego człowieka. Motywacją Autora jest stworzenie systemu, który ułatwi pracę osób analizujących obrazy z monitoringu miejskiego poprzez automatyczne rozpoznawanie i klasyfikację wybranych zachowań/akcji osób na ulicach.

W tym kontekście Autor precyzuje trzy główne cele pracy, którymi są: 1) czasoprzestrzenna lokalizacja i rozpoznawanie w czasie rzeczywistym wielu indywidualnych akcji człowieka na nagraniach z systemów monitoringu ulicznego; 2) utworzenie symbolicznej reprezentacji ruchomych obiektów na podstawie analizy obrazu sekwencji wideo; 3) utworzenie kompleksowego rozwiązania, które pozwoli osobie bez znajomości metod uczenia maszynowego i wizji komputerowej zrealizować proces zdefiniowania zadanej akcji człowieka i włączenia jej w system monitorowania.

Pewne zamieszczenie w opisie dokonań Doktoranta wprowadza sekcja 4.5 poświęcona kontrybucjom Autora rozprawy. Jej główną treścią jest omówienie wybranych prac Autora, przy czym prace te są wieloautorskie, zawierają wspólne dokonania, prawdopodobnie wypracowane w składzie realizującym projekt SAVA, a dodatkowo narracja w pracy prowadzona jest w pierwszej osobie liczby mnogiej. W tym kontekście cel recenzji rozprawy doktorskiej, którym jest wydzielenie i ocena indywidualnych dokonań Doktoranta w dyscyplinie naukowej informatyka techniczna i telekomunikacja, został nieco utrudniony. Przyjęto zatem, że metody opisane w pracach, których Doktorant jest pierwszym autorem i ich rozszerzenia opisane w rozprawie, są głównym osiągnięciem Doktoranta.

Rozdział piąty obejmuje przegląd literatury oraz przegląd aktualnych rozwiązań technicznych w obszarze problematyki objętej rozprawą. Inteligentna analiza obrazu wideo jest bardzo szerokim obszarem badań o wysokiej przydatności praktycznej. Przykłady zastosowań systemów monitorowania, spotykane architektury systemów zostały dokładnie opisane, zaś dyskusja popularnych etapów przetwarzania oraz metod analizy obrazów stanowią wprowadzenie w poszczególne problemy badawcze, w ramach których Doktorant zaproponował autorskie rozwiązania. Przegląd bazuje głównie na pozycjach sprzed kilku lat, gdyż starsze pozycje są dokładniej omówione, zaś publikacje nowsze jedynie reportersko zacytowane. Nie wszystkie wątki są wyczerpująco omówione (np. metody bazujące na sieciach głębokich), ale z uwagi na rozległość dziedziny badań systematyka problemu została przedstawiona poprawnie i klarownie, a podstawowy nurt badań właściwie przedyskutowany.

Głównym podjętym przez Autora zagadnieniem jest wizyjne rozpoznawanie czynności (akcji) wykonywanych przez człowieka. Wśród najpopularniejszych rozwiązań problemu znajdują się metody szacujące pozę człowieka, gdzie kluczowe punkty szkieletu są wykrywane lub śledzone na obrazie i na tej podstawie szacowana jest sekwencja jego póz oraz wynikające z tego działania. Nieco

bardziej ogólna grupa metod bazuje na śledzeniu wizualnych cech poszczególnych obiektów (np. kończyn postaci) pomiędzy obrazami sekwencji wideo. Albo przeprowadzają dla kolejnych klatek proces wykrywania cech charakterystycznych obiektu, określając parametry czasoprzestrzenne śledzonego obiektu, albo poszukują korelacji fragmentów obrazu, co zapewnia dużo większą wydajność procesu śledzenia zmian. To właśnie ta grupa metod stała się przedmiotem badań Doktoranta, głównie z powodu jej uniwersalności, modularności i przydatności w warunkach rzeczywistych, gdzie często mamy do czynienia z niekorzystnymi warunkami oświetleniowymi, słabą jakością obrazu lub częściowym przysłonięciem obiektów. Wśród bezszkieletowych form opisu ruchu (akcji) szczególną uwagę Doktoranta zwróciła metoda opisu czynności za pomocą zbiorów wzorców (sekwencji) ruchu (ang. *Bag of Visual Words*). Przegląd obejmuje również podrozdział poświęcony metodom stosującym głębokie sieci neuronowe (5.2.4). Rozdział ten pozostawia pewien niedosyt, gdyż współczesne rozwiązania, stosujące metody uczenia maszynowego, osiągają bardzo dobre rezultaty, zarówno w ekstrakcji kluczowych cech obrazu jak również w procesach śledzenia, klasyfikacji, czy rozpoznawania obiektów.

Podsumowując przegląd jest dość pobieżny, ale ciekawy. Przedstawia obszary niedoskonałości literaturowych rozwiązań. Pewien niedosyt oceny rozwiązań referencyjnych wynika z braku ilościowej oceny tych rozwiązań, gdyż większość ze sformułowań krytycznych, oceniających wydajność, czy efektywność istniejących metod nie została opatrzona mierzalnymi wielkościami. Odzwierciedlają jednak w pełni stan zagadnienia.

Rozdziały od 6 do 9 przedstawiają kluczowe fragmenty autorskich metod pozwalających na śledzenie aktywności postaci w materiale wideo pochodzącym z monitoringu ulicznego. Rozdział 6 poświęcony jest opisaniu zagadnienia ciągłej kalibracji kamer i śledzenia parametrów ustawienia dynamicznej kamery w czasie rzeczywistym. Kalibracja kamery realizowana jest w oparciu o klasyczne etapy: usunięcie nieliniowych zniekształceń obrazu, będących wynikiem niedoskonałości geometrii optyki oraz obliczenie wewnętrznych i zewnętrznych parametrów kamery na podstawie punktów charakterystycznych obrazu. Pozwala to w efekcie oszacować macierz projekcji obiektów sceny na płaszczyznę obrazu i określić wzajemne położenie oraz orientację kamery i obiektu/-ów sceny. Stosowane metody są w większości znane w literaturze, a nowość dokonań Doktoranta polegała głównie na optymalizacji i przetestowaniu hiper-parametrów wykorzystywanych metod. Według opisu, wyniki eksperymentu pozwoliły wybrać najlepsze spośród testowanych hiper-parametrów, choć nie przedstawiono dyskusji odnośnie uniwersalności dobranych parametrów. Określanie pozycji kamery (m.in. pozycja, orientacja) było drugim etapem kalibracji kamer, która była niezbędna do określenia wzajemnych relacji czasoprzestrzennych kamer i obserwowanych obiektów. Ten etap jest szczególnie istotny, aby prowadzić proces śledzenia zmian w scenie z perspektywy różnych, modyfikujących swoje ustawienia kamer PTZ, umiejscowionych w różnych punktach przestrzeni. Uwspólnianie układu odniesienia (ang. *World Reference Frame*) na podstawie chmury niezmiennych punktów charakterystycznych jest niezwykle przydatne przy tego typu działaniach. Punkty charakterystyczne były definiowane w oparciu o deskryptor SURF, a ich odpowiedniki w poszczególnych ujęciach za pomocą metody losowego konsensusu (RANSAC). Autor zaproponował ciekawe ujęcie problemu, zakładając obecność w kadrze poziomej płaszczyzny podłoża, co usprawniło obliczenia, a w przypadku monitoringu ulicznego jest to jak najbardziej uzasadnione.

Kolejnym etapem, zmierzającym do oceny aktywności ludzi w kadrze jest zadanie oddzielenie tła od poruszających się i nie znanych wcześniej obiektów. W tym zakresie Autor bazował chyba na dość skromnym zestawie referencji, gdyż trudno się zgodzić ze stwierdzeniem, że „tylko kilka prac porusza

ten problem” (str. 83, poz. [14,64,211]). Ostatecznie zastosowano w tym celu metody z biblioteki OpenCV (m.in. MOG). Obiekty ruchome wydzielone z tła pogrupowano z kolei według autorskiego algorytmu w tzw. regiony wykrytego ruchu (ang. *Detected Motion Region - DMR*). Każdy taki rejon był opatrzony pozycją w układzie sceny (WRF).

Rozdział 7 poświęcony jest wydzieleniu z obszaru ruchu (DMR) podobszarów zajmowanych przez aktywność pojedynczych osób (ang. *Separated Motion Region – SMR*). Podstawą do budowania wyizolowanych podobszarów ruchu (SMR) było zdefiniowanie i pogrupowanie ścieżek lokalnych cech (tzw. *Tracks*). Ścieżki były tworzone na podstawie śledzenia kluczowych cech IPAN, zbudowanych w oparciu o zewnętrzny kontur postaci, wyekstrahowany metodą Canny’ego. W oparciu o algorytm węgierski (ang. *Hungarian algorithm*) i opracowaną macierz kosztu, następowało budowanie i ewentualnie agregowanie ścieżek przemieszczania się kluczowych punktów w przestrzeni. Metoda jest ciekawa i perspektywiczna, ale nie wszystkie jej aspekty zostały precyzyjnie wyjaśnione. Przykładowo nie znane są dokładne kwestie stosowania filtra wzdłuż ścieżek cech (str. 104-105) jak również nie dokonano odpowiedniej analizy aktywności ludzi i wynikających z nich założeń odnośnie wartości progowych (np. *threshold* 50 – str. 105). Trudno też ocenić końcową dokładność metody, skoro jak przyznaje Autor na str. 106 metoda nie została zoptymalizowana ze względu na dużą liczbę parametrów metody.

Doktorant słusznie podjął próbę klasteryzacji ścieżek cech kluczowych, aby zbudować na ich podstawie wzorce zachowań poszczególnych postaci. W rozdziale 7.3 przedstawiony jest formalny opis klasteryzacji ścieżek, ale zawiera on wiele empirycznych wskaźników (współczynniki wag *rho* – str. 109), nieprecyzyjnie osadzonych w dziedzinie problemu, jak również chyba nie wszystkie operatory zostały precyzyjnie wytłumaczone, np. operator dodawania lub odejmowania ścieżki T i klastra S , gdy wymiary ich wektorów są różne. Rzutuje to na czytelność rozdziału 7.3.3., który bardzo ogólnie podejmuje problem grupowania hierarchicznego ścieżek. O ile ideowy schemat w sekcji 7.4.1 jest czytelny to trudno wywnioskować z niego parametry metody w warunkach rzeczywistych (sekcja 7.4.2). Analogicznie bardzo ogólnie choć obiecująco brzmią zapewnienia z rozdziału 7.5, które wprawdzie przedstawiają możliwość budowania segmentów ścieżek i ich klasteryzacji, ale trudno na podstawie opisu wywnioskować jak zastosować metodę w ogólności. Analogicznie tajemniczo opisano tworzenie symboli, czyli klastrów segmentów. Ewidentnie opisom formalnym przeprowadzanych operacji łączenia i grupowania przydałby się opis semantyczny aktywności człowieka, które symbole mają reprezentować. Brakuje też ilościowej i jakościowej oceny testów w warunkach rzeczywistych, które nominalnie powinny być zawarte w sekcji 7.8.2. Niejasna jest też skuteczność zarządzania klastrami ścieżek w dziedzinie czasu dla różnych wartości progowych (rys. 7.22), gdyż nie wiadomo jakie są wartości referencyjne. Na jakiej podstawie przyjęto przykładowo wartość progową 30000 ramek dla maksymalnej długości segmentów?

Rozdział 8 poświęcony jest symbolicznej reprezentacji ruchu obiektów. Autor zaproponował formalny opis odległości (dystansu) symboli, ale same symbole nie zostały zdefiniowane. Dodatkowo opis formalny nie jest wystarczająco precyzyjny – brak wyjaśnienia wszystkich niezbędnych symboli i operatorów (np. A_s , B_s , $Assign()$, $average()$, $merge()$, $symbol()$, in.). W opisie rozwiązania bezcenne byłyby schematy i rysunki demonstrujące zasady działania metody. Pomimo niedoskonałości opisu, Doktorant umieścił wyniki testów opracowanego rozwiązania na dwóch referencyjnych zbiorach danych KTH i VMAS₂. Dla selekcji binarnych atrybutów symboli wykorzystany został algorytm CMIM, a do ich klasyfikacji zastosowano klasyfikator SVM. Testy przeprowadzono dla empirycznego zbioru parametrów metod, a wyniki przedstawiono jedynie dla

zadania dwuklasowego rozróżniającego bieg od chodu. Nie wiadomo dlaczego Autor wybrał tylko dwie spośród sześciu (KTH) lub nawet 20 klas zachowań (VMASS₂). Czy dla pozostałych klas testowanych zbiorów danych wyniki metody są porównywalnie dobre? Pewien niedosyt pozostawia brak odniesienia uzyskanych rezultatów do wyników metod referencyjnych.

Rozdział 9 poświęcony jest ocenie rozwiązań potrafiących uczyć się modeli ludzkich zachowań/akcji, realizowanych w czasie rzeczywistym. Autor przeprowadził bardzo staranną analizę popularnych literaturowych zbiorów danych oraz zaproponował ciekawą metodologię oceny i wyboru najlepszej metody przy dużej różnorodności raportowanych w literaturze wyników. Na podstawie analiz wytypowana została metoda opracowana przez F. Shi et al. [240], której wydajność Autor istotnie usprawnił. Usprawnienia bazują głównie na doborze lepszych parametrów metody, aby poprawić czas przetwarzania i wielkość modelu. Wyniki usprawnionej metody nie zostały jednak przedstawione wyczerpująco i nie wiadomo czy wraz ze zwiększeniem wydajności metody zmianie uległa skuteczność samej metody.

Końcowa część pracy (rozdziały 10-11) poświęcone są kolejno opisowi architektury systemu SAVA i opisowi systemów narzędziowych, wspierających system. Pracę kończy rozdział 12 podsumowujący całość osiągnięć i przedstawiający perspektywy dalszych badań oraz możliwości rozwoju systemu monitoringu.

Część trzecia rozprawy to dodatki zawierające listingi kodu, dodatkowe schematy i techniczny manual systemu SAVA.

Przedłożona do oceny rozprawa prezentuje logicznie i chronologicznie najważniejsze rezultaty badań objętych jej tematyką. Mimo, że Doktorantowi nie udało się uniknąć niewielkich potknięć językowych, drobnych nieprecyzyjności formalnych i nie wszystkie komponenty zostały kompleksowo zweryfikowane to zakres i sposób ujęcia prezentowanego materiału pracy oceniam bardzo pozytywnie. Bezsprzecznie należy uznać, że cele pracy zostały osiągnięte, a zakres prac opisanych w rozprawie doktorskiej stanowi jedynie fragment działań Doktoranta objętych pracami w projekcie i problemów, które musiał pokonać Doktorant realizując tak złożony system jakim jest system SAVA.

3. Uzyskane wyniki

Do kluczowych wyników prac i osiągnięć Doktoranta niewątpliwie można zaliczyć:

1. Metodę automatycznego obliczania relacji geometrycznych pomiędzy kamerą a sceną, w której kluczowym elementem jest szacowanie poziomej płaszczyzny podłoża sceny. Dzięki tej metodzie (odnajdywaniu kluczowych punktów podłoża) możliwe jest szacowanie aktualnych pól kamer monitoringu w scenie i w konsekwencji dokładne szacowanie pozycji czasoprzestrzennych śledzonych obiektów.
2. Metodę rozdzielania tła od pierwszoplanowych obiektów (postaci) poruszających się w kadrze, dzięki którym możliwe jest wyznaczenie regionów obejmujących ruchome obiekty (DMR - *Detected Motion Region*).
3. Metodę tworzenia i hierarchizacji ścieżek punktów kluczowych ruchomych obiektów, które mają duży potencjał do definiowania bezszkieletowych, hierarchicznych wzorców ruchu obiektów w obrazie.
4. Optymalizację czasową i wydajnościową opracowanych rozwiązań, które pozwoliły utworzyć spójny i kompletny system monitorowania przestrzeni miejskiej klasy przemysłowej

o modułowej budowie i rozbudowanych funkcjonalnościach – konkurencyjnych względem innych systemów monitoringu, obecnych na rynku.

5. Integrację na poziomie przemysłowym (TRL 9) kluczowych metod i komponentów systemu SAVA, które pozwalają na adaptację i rozbudowę systemu osobom bez doświadczenia informatycznego.

Przedstawiony zakres dokonań Doktoranta należy uznać za istotny i mający wkład do dyscypliny informatyka techniczna i telekomunikacja, w szczególności w zakresie metod analizy cech obrazów i śledzenia ruchomych obiektów w sekwencjach wideo. Należy również podkreślić, że wyniki badań zostały opublikowane w kilkunastu pracach – głównie w tomach pokonferencyjnych renomowanych konferencji międzynarodowych, które zostały wymienione w bibliografii rozprawy doktorskiej i zostały nieco szerzej omówione w sekcji 4.5. Trochę szkoda, że Doktorant zmniejszył ostatnio swoje zainteresowanie tematem badań podjętych w doktoracie, gdyż jak łatwo zauważyć główne publikacje pochodzą z lat 2014-2018. Dalsze utrzymanie efektywności (konkurencyjności) komponentów i metod zaimplementowanych w systemie SAVA wymagałoby dalszego rozwoju metod przykładowo poprzez sięgnięcie do metod uczenia maszynowego. Szerzej ujmując dorobek Doktoranta, w bazie SCOPUS można natomiast znaleźć 44 prace z Jego udziałem. Prace z udziałem Doktoranta zostały zauważone w środowisku, o czym świadczy wysoka liczba cytowań (216 na dzień sporządzania recenzji) oraz współczynnik Hirscha równy 7. Są to bardzo dobre wskaźniki bibliometryczne świadczące o dużo szerszych zainteresowaniach naukowych, niż te podjęte w rozprawie doktorskiej.

4. Uwagi

Lektura rozprawy doktorskiej, która obejmuje tak wiele modułów i wątków związanych z opracowaniem komponentów wizyjnego systemu monitorowania klasy przemysłowej skłania do pytań i refleksji, które w żaden sposób nie podważają zaprezentowanych wyników i osiągnięć, ale świadczą, że przedstawiona do oceny praca jest ciekawa i inspirująca oraz ma potencjał do dalszych badań.

1. Skoro głównym przeznaczeniem wizyjnego systemu monitorowania jest wykrywanie i rozpoznawanie zachowań (akcji) ludzi zarejestrowanych przez kamery wideo to bezcenne wydaje się przedyskutowanie semantyki ruchów ciała, ich trajektorii, czy czasu trwania ruchu z perspektywy parametrów technicznych urządzeń rejestrujących i skomplikowanych warunków środowiskowych. Stanowiłoby to prawdopodobnie dobry punkt wyjścia do przyjmowania różnych założeń i warunków brzegowych (parametrów) dla opracowywanych wzorców ruchu.
2. Opracowane autorsko rozwiązanie (schemat z rysunku 5.7) nie został prawdopodobnie zweryfikowany ilościowo lub wyniki takiej weryfikacji nie zostały przedstawione w sposób klarowny. Rozdziały 6-8 przedstawiają niezależne komponenty metody i jeśli je zweryfikowano to niezależnie od pozostałych elementów. Dogłębna analiza ilościowa i jakościowa poszczególnych komponentów i całego systemu wzmocniłaby istotnie wiarygodność opracowania.
3. W rozprawie (sekcja 4.5) przytoczono szereg wieloautorskich prac naukowych z udziałem Doktoranta, opublikowanych w ramach badań prowadzonych w trakcie tworzenia systemu SAVA. Niestety w całej pracy Doktorant używa liczby mnogiej („my”) i dodatkowo nie określił własnego wkładu w opracowane rozwiązania. Chociaż domyślnie przyjąłem, że elementy opisane w rozprawie są wyłącznie jego autorstwa, to powinno to zostać jednoznacznie wyjaśnione.

4. Stosowany w sekcjach rozdziału 7 zapis formalny nie jest wystarczająco objaśniony. Część z symboli oraz większość operacji logicznych na wektorach opisujących ścieżkę, klaster, symbol, czy grupę wymaga dodefiniowania lub doprecyzowania. Analogiczna sytuacja ma miejsce w formalnych zapisach z rozdziału 8.
5. Opis modyfikacji metody F. Shi et al. [240] nie zawiera analizy ilościowej, a jedynie rozważania optymalizacyjne. Powinno to zostać uzupełnione, gdyż Autor deklaruje na str. 155, że optymalizacji uległa również dokładność metody autorstwa Shi. Warto wyjaśnić czy komponenty opisane w rozdziałach 6-8 zostały wykorzystane w optymalizacji metody z rozdziału 9? Czy opracowane rozwiązania współpracują ze sobą w ramach jednego systemu? Jeśli tak to w jaki sposób?

4.1 Drobne uwagi edycyjne:

1. Umieszczony na stronach 54-55 opis i schemat (rys. 5.7) autorskiego rozwiązania chyba omyłkowo został zakwalifikowany do rozdziału 5 obejmującego przegląd literatury.
2. W ostatnim akapicie na stronie 119 prawdopodobnie zamiast symbolu P w wyrażeniu $D(C,P)$ powinien występować symbol T.
3. Niektóre rozdziały były bardzo krótkie i pozbawione szczegółów (7.8.2; 7.10; 7.12) przez co ich zasadność w obecnej postaci jest trudna do wyjaśnienia.
4. Nie wszystkie rysunki mają odwołania w tekście pracy (np. 7.20, 7.21, 7.22) przez co ich odmiejszczenie utrudnia lekturę niektórych fragmentów rozprawy.
5. W opisie do rys. 6.5 Autor wspomina o niebieskim prostokącie z czerwoną obwódką, jednak na rysunku nie widać żadnej czerwonej obwódki.
6. Numeracja dodatków do rozprawy jest nie do końca spójna i zrozumiała. Wprowadza to przykładowo zamieszanie w odwoływaniach do poszczególnych jego sekcji – np. odwołanie do sekcji 3.6 dodatku na stronie 43 pracy, nie zawiera schematów kluczowych bloków metody BoVW, gdyż są one na stronach 261-268.
7. W pracy można znaleźć nieliczne błędy edytorskie i językowe, np.: andr (str. 89); thee (str. 104); z_{min} (str. 114); used by used by (str. 133); original, original (str. 156);

5. Podsumowanie

Przytoczone przeze mnie niewątpliwe osiągnięcia Doktoranta, nawet w świetle zgłoszonych uwag, pozwalają stwierdzić, że Pan mgr inż. Marek Kulbacki wykazał się właściwą wiedzą z zakresu tworzenia, adaptacji i optymalizacji metod wykrywania i śledzenia zachowań postaci w sekwencjach obrazu wideo, a także opanował i właściwie posługuje się warsztatem naukowym. Zadał też o to, aby wyniki jego prac były odpowiednio udokumentowane i spopularyzowane, o czym świadczą liczne i zauważone przez środowisko publikacje naukowe. Uważam, że przedłożona do oceny rozprawa doktorska Pana mgr inż. Marka Kulbackiego zawiera oryginalne i kompleksowe rozwiązania ważnego i aktualnego problemu badawczego, jest dowodem na wysoką ogólną wiedzę teoretyczną i umiejętności prowadzenia przez Doktoranta prac naukowych. Tym samym **rozprawa** moim zdaniem **spełnia z wyraźnym nadmiarem** wymogi Ustawy z dnia 14.3.2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki, z późniejszymi zmianami. Osiągnięcia Doktoranta wnoszą bardzo wyraźny i istotny wkład naukowy w dyscyplinę naukową informatyka techniczna i telekomunikacja. Wnioskuje o dopuszczenie Doktoranta do dalszych etapów przewodu doktorskiego.

Adam Wojewechowski 7

