20 maja 2019 r.

prof. dr hab. inż. Paweł Strumiłło
Instytut Elektroniki
Politechniki Łódzkiej

RECENZJA ROZPRAWY DOKTORSKIEJ
DLA WYDZIAŁU ELEKTRONIKI I TELEKOMUNIKACJI
POLITECHNIKI POZNAŃSKIEJ

Tytuł rozprawy: *Hybrid techniques of depth map estimation and their application in three-dimensional video systems (Hybrydowe techniki wyznaczania map głębi i ich wykorzystanie w systemach obrazu trójwymiarowego)*

Autor rozprawy: mgr inż. Maciej Kurc

Promotor: prof. dr hab. inż. Marek Domański

I. Jakie zagadnienie naukowe jest rozpatrzone w pracy /teza rozprawy/ i czy zostało ono dostatecznie jasno sformułowane przez autora?

Przedłożona do recenzji rozprawa dotyczy metod obliczeniowych związanych z wyznaczaniem tzw. map głębi otoczenia. Mapy takie umożliwiają odwzorowanie trójwymiarowej struktury obrazowanej sceny i mają wiele zastosowań, m.in. w nawigacji robotów, medycynie, systemach wspomagania osób niewidomych, a także systemach tzw. telewizji swobodnego widzenia. W szczególności w tym ostatnim zastosowaniu jakość uzyskiwanych map głębi o dobrej jakości warunkuje uzyskanie dobrej jakości obrazów (widoków) sceny wyznaczanych dla arbitralne ustalonej wirtualnej pozycji obserwatora. Techniki wyznaczania map głębi można podzielić na metody pasywne (m.in. wykorzystujące stereowizję) oraz techniki aktywne polegające na emisji energii i rejestracji czasu powrotu sygnału odbitego od obiektów sceny. Metody pasywne wymagają dużej mocy obliczeniowej i charakteryzują się małą dokładnością odwzorowania głębi. Obie te właściwości stanowią istotne ograniczenie w systemach telewizji wielowidokowej, w której wymaga się zarówno rekonstrukcji obrazów z szybkością standardu telewizyjnego oraz z dużą dokładnością odwzorowania zależności geometrycznych sceny. Ograniczenia te wskazał Autor pracy i podjął się opracowania alternatywnej metody, którą nazwał metodą hybrydową do wyznaczania sekwencji map głębi z zastosowaniem techniki aktywnej oraz synchronizowanych z nią sekwencji obrazów sceny. Autor rozprawy zastosował tzw. kamerę czasu przelotu (*Time of Flight –ToF*) umożliwiającą odwzorowanie map głębi z większa dokładnością w porównaniu z mapami wyznaczanymi ze stereowizji. Zastosowanie kamer głębi ToF wiąże się z większym zużyciem energii (emisja energii świetlnej), wymaga

jednak o wiele mniej obliczeń w porównaniu ze stereowizją, w której każda wartość (punkt) w mapie głębi wiąże się z poszukiwaniem minimum pewnej zadanej funkcji kosztu.

Autor sformułował i wykazał w pracy następujące tezy (zapisane w wersji angielskiej podanej przez Autora rozprawy i przetłumaczone na język polski przez recenzenta):

T1. *A hybrid system that estimates depth both from video data and ToF camera output can be relatively easily built and programmed. Such a system allows to obtain higher quality depth maps when using additional information from ToF camera(s) instead of only stereo correspondence* (System hybrydowy estymacji głębi na podstawie danych wideo i danych z kamery ToF może być stosunkowo łatwo zbudowany i zaprogramowany. System taki umożliwia wyznaczanie map głębi o lepszej jakości przy wykorzystaniu dodatkowej informacji z kamer ToF zamiast informacji pozyskanej wyłącznie ze stereowizji.)

T2. *Depth map inter-view consistency improvement algorithm that increases depth map quality and increases the compression efficiency of compression algorithms which exploit inter-view relations of depth maps* (Algorytm poprawy międzywidokowej spójności mapy głębi zwiększa jakość mapy głębi oraz zwiększa wydajność algorytmów kompresji wykorzystujących zależności map głębi w różnych widokach.)

Do ww. tez rozprawy zgłaszam następujące uwagi:

1. Pierwsze zdanie w tezie pierwszej jest zbyt ogólne. Przymiotnik „łatwy" nie jest techniczną wielkością mierzalną. Teza ta mogłaby składać się tylko z jednego zdania sformułowanego np. w brzmieniu: *A hybrid system for estimating depth maps both from video data and ToF camera output allows to obtain higher quality depth maps than the maps computed from stereo correspondence only.*
2. W drugiej tezie jak można przypuszczać Autor pomyłkowo użył zaimka *that,* który nie powinien w tej tezie wystąpić. W przeciwnym przypadku brakuje w tej tezie orzeczenia.

Pomimo zauważonych ww. usterek w tezach pracy cel badawczy został w pracy jasno sformułowany. Polega on na opracowaniu metody wykorzystującej dane wideo oraz mapy głębi, która umożliwi poprawę skuteczności kompresji w telewizji wielowidokowej oraz jakości rekonstruowanych widoków.

II. Czy tematyka rozprawy jest aktualna lub dostatecznie ważna?

Tematyka rozprawy jest bardzo aktualna i dotyczy dynamicznie rozwijających systemów telewizji swobodnego puntu widzenia (FTV – *Free-Viewpoint Television*). Autor skoncentrował się w pracy na kluczowych problemach obliczeniowych związanych z technikami kompresji sekwencji wielowidokowych. Dla osiągnięcia tego celu podjął się zadań badawczych prowadzących do poprawy jakości map głębi oraz ich tzw. spójności czasowej (*time-consistency*) i spójności przestrzennej (*inter-view consistency*), które warunkują uzyskiwanie dobrych parametrów kompresji i jakości syntezowanych widoków.

III. Czy autor rozwiązał postawiony problem i czy użył to do tego właściwych metod?

Zgodnie z postawionymi tezami pracy Autor sformułował następujące szczegółowe cele badawcze pracy:

C1. Estymacja parametrów systemu wielokamerowego, rektyfikacja obrazów oraz redukcja zakłóceń w mapie głębi.
C2. Korekcja pomiarów głębi uzyskiwanych z zastosowaniem kamery ToF.
C3. Wzajemna synchronizacja kamer wideo i kamer głębi ToF.
C4. Fuzja obrazów rejestrowanych z kamer wideo i kamer głębi ToF.
C5. Poprawa jakości map głębi na podstawie poprawy spójności przestrzennej widoków.

Ad. C1. Estymacja parametrów systemu wielokamerowego, rektyfikacja obrazów oraz redukcja zakłóceń w mapie głębi: Ten cel rozprawy można uznać za pomocnicze ale zasadnicze zadanie projektowe. Rejestracja obrazów z wielu kamer, a następnie rekonstrukcja trójwymiarowej geometrii sceny wymaga identyfikacji wzajemnych geometrycznych położeń kamer, zadanie to określa się mianem estymacji parametrów zewnętrznych kamer tj. obliczenia wektorów translacji i obrotu, definiujących wzajemne przemieszczanie kamer. Autor wykorzystuje to tego celu znany algorytm Zhanga, w którym stosuje tablice kalibracyjne. Wprowadza modyfikacje do tego algorytmu polegające na wykorzystaniu danych o głębi uzyskiwanych z kamer ToF. Po przeprowadzeniu rektyfikacji obrazów, tj. wprowadzenia estymowanych parametrów identyfikujących geometrie kamer do układu liniowego, Autor uzyskał ponad dwukrotną redukcję błędu średniokwadratowego estymacji położenia punktów obliczanego miarą odległości Euklidesowej dla kamer głębi ToF.

Ad. C2. Korekcja pomiarów głębi uzyskiwanych z zastosowaniem kamery ToF. W pierwszym kroku Autor przeprowadza korekcję systematycznych błędów wyznaczania głębi w kamerze ToF. Kamera ToF mierzy odległości do elementów sceny trójwymiarowej i wg firmowego (niedostępnego użytkownikowi) algorytmu oblicza wartości głębi. Autor pokazuje, że błędy wyznaczania głębi można redukować, na podstawie danych o parametrach wewnętrznych kamery (ogniskowej i współrzędne środka optycznego kamery). Z zastosowaniem zaproponowanego algorytmu zarówno błąd maksymalny jak i błąd średniokwadratowy zostały kilkukrotnie zmniejszone odpowiednio do wartości E_{RMS}<4mm i E_{MAX}<10mm. Przy czym zauważa się tym mniejszą skuteczność redukcji błędów im obiekty są położone dalej od osi optycznej kamery (wg miary kątowej). W drugim etapie korekcji pomiarów głębi Autor skoncentrował się na redukcji błędów związanych z zakłóceniami obrazów głębi. Zaproponował przestrzenną filtrację adaptacyjną zachowującą krawędzie natomiast skuteczniej filtrujących obszary o jednolitych wartościach głębi zawierających zakłócenia. Zastosował również filtrację dla sekwencji obrazów głębi (filtr rekursywny), w których filtracji poddawany są jedynie obszary nieruchome. Na przykładowych mapach głębi pokazał wyniki zastosowanej filtracji i poprawę jakości map głębi. Nie podał jednak wartości ilościowych dla efektów zastosowanych filtrów.

Ad. C3. Wzajemna synchronizacja kamer wideo i kamer głębi ToF. To osiągniecie pracy jest techniczne co nie znaczy, że umniejsza to wartości tego osiągnięcia. Zapewnienie synchronizacji kamer (synchroniczności zbieranych obrazów sekwencji wideo) jest konieczne do utrzymania spójności czasowej całego procesu kompresji obrazów wielowidokowych (co Autor klarownie wyjaśnia we wstępie do rozdziału czwartego i pokazuje przykłady błędnie wyznaczanych map głębi przy braku synchronizacji kamer). Autor zbudował układ elektroniczny do synchronizacji kamer z wykorzystaniem układów programowalnych FPGA, jego budowę i skuteczne działanie wyjaśnił w dodatku A pracy.

Ad. C4. Fuzja obrazów rejestrowanych z kamer wideo i kamer głębi ToF w celu uzyskania poprawy jakości map głębi. W tym kluczowym wątku badawczym Autor podjął się zadania łączenia obrazów rejestrowanych z kamer wideo oraz z kamer ToF umieszczonych w układzie liniowym, ale różnej liczbie i konfiguracjach przestrzennych oznaczanych przez Autora kodami 3T+2D (trzy kamery wideo i dwie kamery głębi), 5T+2D, 6T+3D. Zastosował dwie metody oceny jakości map głębi, tj. na podstawie jakości syntezy widoku wirtualnego oraz na podstawie bezpośredniego porównania obliczonej mapy głębi i „prawdziwej" mapy głębi wyznaczonej inną dokładną metodą odniesienia. Jako materiał badawczy w pierwszej metodzie weryfikacji jakości map głębi wykorzystał własne wielowidokowe rejestracje scen trójwymiarowych, natomiast w drugiej metodzie wykorzystał dane z bazy Middlebury, zawierającej obrazy wielowidokowe oraz mapy głębi i te mapy głębi wykorzystał jako symulowane mapy głębi uzyskane z kamery ToF. Przedstawił bardzo obszerne zestawienia tabelaryczne z wynikami oceny jakości map głębi (wg miar podanych w pracach Scharsteina oraz miary PSNR (*Peak Signal to Noise Ratio*)). W zaproponowanej metodzie łączenia (fuzji) obrazów z kamer wideo i kamer głębi stosował regułę wyznaczania głębi sterowaną miarą pewności dla obliczanych wartości głębi ze stereowizji. Uzyskana poprawa jakości mapy głębi jest jednoznacznie pozytywna dla miar Scharsteina (liczba tzw. „złych pixeli"), natomiast wniosku takiego nie można wysnuć dla miary PSNR. Cennym spostrzeżeniem jest fakt, że poprawę jakości map głębi uzyskano głównie dla obszarów bez wyraźnej tekstury, tj. obszarów, dla których głębia ze stereowizji jest często wyznaczana błędnie. Osiągnięcie tego celu oznacza wykazanie tezy pierwszej pracy.

Ad. C5. Poprawa jakości map głębi na podstawie poprawy spójności przestrzennej widoków. Celem niniejszego zadania badawczego jest opracowanie metody poprawy spójności między widokowej mapy głębi, w której nie jest konieczna jej re-estymacja wiążąca się ze znacznym kosztem obliczeniowym. Algorytm ten polega na dodatkowym przetwarzaniu map głębi i wzajemnego wykorzystywania danych z map głębi w sąsiednich widokach. Zaproponowany algorytm jest iteracyjny i wykorzystuje filtry medianowe (jeden z filtrów jest adaptacyjny). Działanie zaproponowanego algorytmu poddał weryfikacji eksperymentalnej dla testowych sekwencji wielowidokowych (część z nich jest autorstwa Doktoranta). Uzyskał znaczącą poprawę miedzy widokowej spójności map głębi. Pokazał, że już po trzech iteracjach algorytmu miara niespójności osiąga małą wartość. Ważnym efektem tego algorytmu jest poprawa stopnia kompresji sekwencji wielowidokowych dla kodeków stosowanych do kompresji w systemach

telewizji wielowidokowej (redukcja przepływności bitowej o ok. 5%). Wyniki te dowodzą tezy drugiej pracy.

IV. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek autora? Jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanych przez literaturę światową?

Rozprawa koncentruje się na nowoczesnym i ważnym dziale techniki związanym z przetwarzaniem i kompresją obrazów oraz kompresji sekwencji obrazów wielowidokowych stanowiących techniczną podstawę systemów telewizji swobodnego punktu widzenia (*FreeViewpoint Television – FTV*). Jest to dział techniki będący obszarem zainteresowania badawczego wielu ośrodków naukowych na świecie dysponujących dużymi budżetami na badania. Osiągnięcia badawcze Autora rozprawy są na poziomie światowym, o czym świadczą Jego wartościowe publikacja i przyznane patenty. W rozprawie Autor pokazał, że posiada jak i potrafi skutecznie korzystać z nabytej interdyscyplinarnej wiedzy technicznej z dyscyplin telekomunikacji, informatyki i elektroniki. Zaproponował oryginalne algorytmy oraz układy elektroniczne prowadzące do poprawy zarówno jakości syntezowanych widoków jak i skuteczności kompresji sekwencji obrazów wielowidokowych. Wskazał kluczowe elementy systemu kompresji sekwencji obrazów wielowidokowych, dla których zaproponował istotne i oryginalne modyfikacje z zastosowaniem kamer głębi ToF. Zaproponował i poddał weryfikacji oryginalne metody rektyfikacji obrazów i redukcji zakłóceń w obrazach rejestrowanych przez kamery ToF. Zbudował złożony układ elektroniczny do synchronizacji kamer wideo i kamer głębi ToF oraz pokazał, ze poprawa spójności przestrzennej mag głębi umożliwia uzyskiwanie poprawy przepływności bitowej sekwencji obrazów wielowidokowych.

V. Czy rozprawa świadczy o dostatecznej wiedzy autora i znajomości współczesnej literatury z dyscypliny naukowej, której dotyczy?

Autor biegle porusza się w tematyce rozprawy związanej z przetwarzaniem obrazów cyfrowych i kompresją obrazów sekwencji wielowidokowych. W okresie pracy nad rozprawą był członkiem zespołu, który ma renomę międzynarodową w tej tematyce badawczej. Zamieszczony w pracy wykaz literatury liczy ponad 200 pozycji literatury nie licząc pozycji współautorskich Autora rozprawy, na które składają się: jedna publikacja w czasopiśmie naukowym z listy JCR, 5 publikacji w czasopismach krajowych i 19 referatów na konferencjach międzynarodowych.

VI. Jakie są słabe strony rozprawy i jej główne wady?

Strona formalna rozprawy jest bez zastrzeżeń. Praca jest napisana dobrym językiem angielskim i opracowana starannie pod względem edycyjnym i graficznym. Autor wyróżnia pogrubioną czcionką najistotniejsze wnioski i osiągnięcia pracy.

Punkty dyskusyjne, na które chciałbym zwrócić uwagę w pracy są następujące:

1) Teza druga rozprawy wymaga korekty gramatycznej (zaznaczonego w pierwszym punkcie niemniejszej recenzji).

2) W sekcji pracy opisującej poprawę skuteczności algorytmu estymacji parametrów zewnętrznych dla konkretnych modeli kamer ToF Autor mógłby przeprowadzić bardziej formalną analizę matematyczną zagadnienia.
3) W rozdziale 3. Autor stosuje filtr o nieskończonej odpowiedzi impulsowej do filtracji obszarów map głębi nie będących w ruchu. Dlaczego zastosowano filtr o nieskończonej odpowiedzi impulsowej, którego odpowiedź na filtrowane sygnały (jak sama nazwa wskazuje) będzie trwała przez wiele następnych obrazów w sekwencji wideo? Może to wprowadzać dodatkowe zakłócenia do map głębi.
4) W rozdziale 3. pokazano poprawę wyglądu map głębi tylko na jednym przykładzie. Nie zaproponowano miar do oceny ilościowej poprawy parametrów map głębi (miarę taką wprowadzono dopiero w rozdziale 6., ale w innym kontekście).
5) W rozdziale 6. zaproponowano algorytm poprawy spójności międzywidokowej map głębi i pokazano znaczenie tego algorytmu dla poprawy przepływności bitowej kodowanych sekwencji. Nie przeprowadzono natomiast szacunkowej oceny kosztu obliczeniowego tego algorytmu, co jak sugeruje Autor powinno być podstawowa zaletą proponowanego rozwiązania.

Wymienione uwagi nie podważają istoty osiągniecia badawczego Autora, a głównie dotyczą niedostatków metodycznych w prezentacji rozprawy.

Autor pracy nie ustrzegł się pewnej liczby zdań nieścisłych lub kolokwialnych, dla przykładu (jest szereg innych):

- str. 63 – „Measured distance data provided by ToF camera is noisy." – zakłóceniu może podlegać sygnał lub obraz ale nie „mierzona odległość".
- Str. 90 – „The synchronisation converter must be aware of the time …" – personifikacja

VII. Wniosek końcowy

Przedłożona do oceny rozprawa doktorska jest wartościowym opracowaniem wnoszącym oryginalny wkład do dyscypliny telekomunikacja w zakresie kompresji sekwencji obrazów wielowidokowych. Dla wykazania tez Autor rozprawy zaproponował oryginalne algorytmy jak i zbudował złożone układy elektroniczne. Postawionych oraz dowiódł na drodze eksperymentalnej i pokazał ilościowe miary poprawy jakości map głębi i przepływności bitowej w systemach telewizyjnych swobodnego punktu widzenia.

Biorąc pod uwagę osiągniecia badawcze Autora rozprawy, bogaty dorobek publikacyjny (jeden artykuł w czasopiśmie naukowym z listy JCR, pięć artykułów w czasopismach krajowych oraz duża liczba referatów konferencyjnych na konferencjach międzynarodowych), a także współautorstwo sześciu wniosków patentowych (dwóch przyznanych) ściśle związanych z tematyką rozprawy **zgłaszam rozprawę doktorską mgra Macieja Kurca do wyróżnienia.**

Przedłożona do recenzji rozprawa spełnia z nadmiarem wymagania Ustawy o stopniach naukowych i tytule naukowym z 14 marca 2003 roku, Dziennik Ustaw Nr 65, poz. 595 odnośnie stopnia doktora nauk technicznych. Wnioskuje o jej dopuszczenie do publicznej obrony.

P. Skulich