prof. dr hab. inż. Andrzej Kasiński
Politechnika Poznańska, Wydział Elektryczny
Instytut Automatyki i Inżynierii Informatycznej

Poznań, 15 listopada 2018 r.

RECENZJA ROZPRAWY DOKTORSKIEJ
DLA RADY WYDZIAŁU ELEKTRONIKI i TELEKOMUNIKACJI
POLITECHNIKI POZNAŃSKIEJ

Tytuł rozprawy: **Synteza widoków wirtualnych w rzadkich systemach wielokamerowych dla zastosowań w swobodnej nawigacji.**

Autor rozprawy: **mgr inż. Adrian Dziembowski**

1. Obszar problemowy rozprawy

Przedmiotem rozprawy jest autorska metoda generowania syntetycznych widoków wirtualnych w wielokamerowym systemie wizyjnym (telewizji cyfrowej), celem stworzenia możliwości swobodnej, wirtualnej nawigacji widza, tj. wirtualnej zmiany jego punktu widzenia względem sceny obserwowanej przy użyciu rzeczywistego, wielokamerowego systemu wizyjnego. Istotną charakterystyką rozważanego systemu wielokamerowego jest w miarę swobodne (rzadkie) rozmieszczenie kamer rzeczywistych w otoczeniu obserwowanej sceny. Ważnym aspektem proponowanej metody syntezy widoku wirtualnego jest docelowa konieczność przeprowadzania jej w trybie czasu rzeczywistego, w którym widoki wirtualne dla wybieranych punktów obserwacji sceny byłyby generowane w tempie zbliżonym do tempa generowania ramek obrazowych w strumieniu wizyjnym kamery rzeczywistej.

Problematyka rozprawy jest kluczowa dla wprowadzenia do powszechnej praktyki systemów telewizji cyfrowej z funkcją swobodnego wyboru punktu obserwacji sceny przez telewidza (tzw. telewizji swobodnego dostępu) i z tego względu jest przedmiotem intensywnych prac badawczych w skali globalnej oraz działań standaryzacyjnych odpowiednich komitetów IUT, ISO/IEC, w szczególności grupy MPEG. Autor rozprawy aktywnie uczestniczy w tych pracach. Znaczenie tej problematyki jest jednak znacznie szersze, gdyż dotyczy innych aplikacji wykorzystujących wielokamerowe systemy wizyjne do wspomagania nawigacji w przestrzeni i planowania trajektorii obiektów np. w robotyce, czy aeronautyce.

Celem zaproponowanej w rozprawie metody jest synteza hipotetycznego widoku w wirtualnej kamerze na podstawie widoków i map głębi sceny, uzyskanych w kamerach rzeczywistych rozmieszczonych (stosunkowo rzadko i niekoniecznie linearnie) w pobliżu wybranego, wirtualnego punktu obserwacji. Podstawą metody syntezy obrazu jest operacja rzutowania rzeczywistej informacji obrazowej i skojarzonej z nią tzw. mapy głębi sceny do układu współrzędnych obrazowych kamery wirtualnej. Rzutowanie, realizowane algorytmicznie, wykorzystuje macierze transformacji wyprowadzone z tzw. pełnych macierzowych modeli kamer realnych i kamery wirtualnej. Efektem rzutowania jest występowania w syntetyzowanym obrazie licznych artefaktów pogarszających jego jakość. Autor rozprawy przedstawia własne sposoby usuwania lub minimalizowania obecności tych efektów, uwzględniając zarazem ich przydatność z punktu widzenia implementacji metody w systemach działających w trybie czasu rzeczywistego.

Teza rozprawy została sformułowania w następujący sposób:

„Możliwe jest opracowanie wydajnej metody syntezy widoków wirtualnych dobrej jakości poprzez wykorzystanie dalszych widoków rzeczywistych, filtrację krawędzi w widokach rzeczywistych i odpowiadających im mapach głębi oraz korekcję niespójności barwnej widoku wirtualnego."

Teza ma charakter cokolwiek nieostry, poprzez twierdzenie jedynie o możliwości uzyskania spodziewanego wyniku oraz relatywny charakter takich określeń jak

„wydajna" czy „dobra jakość". Jednocześnie już w jej treści wskazane są środki jakie zostaną zastosowane dla poprawy jakości obrazu syntetycznego (co jest samo w sobie cenne). Nieostre kwalifikatory występujące w sformułowaniu tezy, zostały skonkretyzowane w toku rozprawy. Przy takiej postaci tezy nie ma innej drogi dla wykazania jej prawdziwości jak eksperymentalna weryfikacja i taką drogę przyjął Autor.

2. Ocena merytoryczna rozprawy.

W tym punkcie dokonam oceny następujących elementów rozprawy: oceny przedmiotowego stanu badań dokonanej przez Autora, przyjętej metodologii badań, przyjętych założeń proponowanej metody, proponowanych sposobów redukcji zakłóceń obrazu syntetycznego, stopnia realizacji celów pracy oraz uproszczeń metody w związku z dążeniem do jej implementacji w systemie czasu rzeczywistego.

Przegląd stanu badań przeprowadzony w rozdziale 2 rozprawy cechuje się starannością i aktualnością, Autor przeanalizował 215 publikacji źródłowych z ostatnich lat, w większości ściśle związanych z problematyką rozprawy. Prawidłowo zidentyfikował ośrodki prowadzące prace bliskie przedmiotowi rozprawy ze względu na cele i zastosowane metody. Nie bez znaczenia był tu aktywny udział Autora rozprawy w latach 2014 – 2018, w pracach grupy MPEG, potwierdzony współautorstwem 20 opublikowanych dokumentów. Przeprowadzona analiza konkurencyjnych prac jest dogłębna i precyzyjna. Ten aspekt rozprawy oceniam bardzo wysoko.

Prace nad telewizją swobodnego dostępu prowadzone są w skali globalnej i są przedmiotem swoistej koordynacji poprzez grupy robocze (w aspekcie ewentualnej standaryzacji systemów) oraz corocznie organizowane konferencje międzynarodowe (w aspekcie przepływu idei). Źródła trudności występujących przy implementacji efektywnych systemów wielokamerowych są zidentyfikowane i dobrze rozpoznane. W szczególności są one związane z występowaniem specyficznych zakłóceń na

obrazie syntetycznym (deficytów i zniekształceń informacji) w porównaniu z obrazem rzeczywistym, zarejestrowanym w tym samym punkcie obserwacji i tej samej orientacji obu kamer: wirtualnej i rzeczywistej. Zakłócenia te immanentnie wynikają już z samych podstaw metody syntezy. Kłopotliwa obecność artefaktów jest funkcją konfiguracji kamer rzeczywistych systemu wielokamerowego, przy czym założenie o rzadkim rozmieszczeniu kamer sprawia, że ich eliminacja staje się szczególnie trudna. Podjęte badania mają na celu zminimalizowanie niepożądanych efektów syntezy i ich wpływu na wynik, którym jest jak najwyższej jakości obraz wirtualny.

Zazwyczaj alternatywne, proponowane metody są standardowo porównywane ze sobą ze względu na jakość uzyskanego obrazu wynikowego w oparciu o tzw. wskaźniki obiektywne (PSNR, SSIM) oraz subiektywną ocenę grupy ekspertów (zgodnie z metodologią oceny przyjętą w ramach grupy MPEG). Materiałem, na którym jest badana efektywność proponowanych ulepszeń metody syntezy są zróżnicowane co do treści, wybrane (standardowe, benchmarkowe) sekwencje video, w tym dwie sekwencje syntetyczne o dostępnej perfekcyjnej mapie głębi (grafika renderowana). Celem porównania efektów kolejnych ulepszeń metody syntezy obrazu wirtualnego dostępne jest również tzw. oprogramowanie referencyjne VSRS (View Synthesis Reference Software), umożliwiające ocenę rezultatów syntezy przy różnych założeniach odnośnie parametrów i konfiguracji systemu.

Autor rozprawy zastosował tę powszechnie przyjętą metodologię do oceny jakościowej rezultatów proponowanej przez siebie rozszerzonej metody syntezy. Dla poprawy reprezentatywności wyników wzbogacił jednak zbiór sekwencji testowych o sekwencje zarejestrowane przez grupę, w ramach której prowadzi badania. Nie było to bez znaczenia, gdyż pomogło uchwycić wpływ kontekstu obrazowego na skuteczność syntezy obrazu wirtualnego.

Istotną innowację stanowiło przyjęcie założenia o rzadkim i swobodnym rozmieszczeniu kamer w otoczeniu sceny, przy jednoczesnej propozycji wykorzystania do syntezy widoku wirtualnego informacji z większej liczby kamer rzeczywistych niż tylko para kamer najbliżej usytuowanych względem wirtualnego

punktu obserwacyjnego. Wykorzystanie tzw. kamer dalszych powoduje, że artefakty w obrazie syntetycznym stają się trudniejsze do usunięcia, zwłaszcza w związku z tzw. efektami odsłaniania, czy zróżnicowanymi charakterystykami foto- i kolorymetrycznymi kamer rzeczywistych przy nie lambertowskim oświetleniu sceny.

Opracowana przez Autora metoda MSV (Multi View Synthesis), obok uwzględnienia dodatkowo innych niż sąsiednie widoków rzeczywistych, wprowadza dodatkowe filtry obrazu syntetyzowanego i skojarzonej z nim mapy głębi, dopuszcza możliwość sztucznego powiększania rozdzielczości obrazów wejściowych i ich map głębi oraz zapewnia korekcję niespójności barw w obrazie syntetycznym. Szczególną uwagę zwrócono na usuwanie pozornych krawędzi w obrazie wirtualnym wynikających z niskiej jakości obrazów wejściowych (rzeczywistych). Ulepszono samą procedurę rzutowania poprzez nałożenie warunku ciągłości obszarów wirtualnych odpowiadających jednorodnym triangulacjom na obrazach źródłowych. Zaproponowano adaptacyjną oraz globalną technikę korekcji niespójności barwowych w obrazach źródłowych. Są to oryginalne elementy w koncepcji proponowanej metody syntezy.

Przeprowadzono liczne i czasochłonne eksperymenty obliczeniowe celem sprawdzenia wpływu zaproponowanych poszczególnych ulepszeń na jakość obrazu syntetycznego. Ich protokoły i wyniki starannie udokumentowano w dodatku do rozprawy. Porównań dokonywano pomiędzy widokami generowanymi metodą autorską, a widokami zsyntetyzowanymi przy użyciu VSRS, przy użyciu obiektywnych i subiektywnych miar jakości. Dzięki wykorzystaniu dalszych widoków w procesie wypełniania obszarów odsłoniętych na obrazie wirtualnym uzyskano w przypadku sekwencji testowych poprawę jakości obrazu syntetycznego średnio o 0.41 dB. Średnia poprawa wartości wskaźnika PSNR dzięki zastosowaniu zaproponowanych modyfikacji wyniosła 1.2 dB w stosunku do obrazów wirtualnych zsyntetyzowanych przy użyciu oprogramowania referencyjnego, Potwierdza to przewagę metody MSV nad innymi znanymi metodami i zarazem potwierdza tezę rozprawy w aspekcie jakościowym. Pozostałe proponowane modyfikacje procedury

syntezy miały (w kategoriach ilościowej oceny jakości obrazu) pozytywny, lecz stosunkowo niewielki wpływ na jakości syntezy. Ich pozytywny wpływ był silniejszy z punktu widzenia oceny jakościowej (subiektywnej). Ostatnia uwaga dotyczy w szczególności eliminacji/redukcji takich artefaktów jak krawędzie pozorne oraz niespójności barwowe obszarów.

Jest charakterystyczne, że w wielu pracach nad poprawieniem jakości syntezy wielowidokowej nie podejmuje się w sposób jawny aspektu praktycznej złożoności obliczeniowej proponowanych metod. Od kwestii tej abstrahuje np. VSRS jako system otwarty, niezależny od platformy obliczeniowej. Pod tym względem Autor rozprawy pozytywnie się wyróżnia. Po zbadaniu wpływu zaproponowanych oryginalnych rozwiązań, ukierunkowanych na poprawę jakości obrazu wirtualnego, Autor podjął analizę kosztów obliczeniowych poszczególnych elementów (etapów obliczeniowych) proponowanej metody syntezy. W jej wyniku zaproponował szereg kompromisów pomiędzy nieuniknioną (acz ograniczoną) utratą jakości obrazu syntetycznego, wynikającą z uproszczeń własnej metody MSV do wersji rt-MSV, a redukcją kosztu obliczeniowego jej implementacji. W rezultacie, przy wykorzystaniu klasycznej architektury CPU, przetwarzania częściowo 4-wątkowego na stosunkowo mocnej platformie (i7-4770, 3.4 GHz), udało mu się ograniczyć czas generowania ramek wirtualnych do 38 ms, średnio w zbiorze sekwencji testowych, a więc spełnić wymagania pracy w trybie czasu rzeczywistego. Dla sekwencji obrazów o rozdzielczości 1920x1080 czas ten wyniósł średnio ok. 50 ms, bez istotnej utraty jakości obrazu. W tym aspekcie można również uznać tezę rozprawy za potwierdzoną. Autor zasygnalizował dalszą możliwość przyspieszenia procesu syntezy pojedynczej ramki na drodze szerszego wykorzystania przetwarzania wielowątkowego w algorytmie syntezy (w przypadku tradycyjnych procesorów), albo przetwarzania masywnie równoległego z wykorzystaniem GPU.

Biorąc pod uwagę docelowo masowy charakter użytkowania systemów telewizji 3D swobodnego dostępu oraz ujawnioną w trakcie badań zależność jakości rezultatów syntezy widoków wirtualnych od kontekstu obrazowego, warte rozważenia wydaje się

użycie metod głębokiego uczenia maszynowego (DML) do ewentualnego wygenerowania architektur szybkich, nisko kosztowych, sprzętowych syntetyzatorów obrazu wirtualnego.

3. Ocena strony formalnej i uwagi szczegółowe.

Rozprawa przedstawiona w języku polskim, obejmująca 182 str. tekstu plus 84 str. dodatków, w których przedstawiono szczegółowe protokoły z przeprowadzonych obliczeń, składa się z 13 punktów/rozdziałów, w tym wprowadzenia, podsumowania i bibliografii. Jest zredagowana bardzo przejrzyście i starannie. Zawiera spis używanych w dalszej treści symboli i skrótów oraz wykaz stosowanych definicji. Podkreślić należy zwarte i eleganckie przedstawienie podstaw metody syntezy obrazu wirtualnego. Z punktu widzenia recenzenta, użyteczne jest zamieszczone na wstępie rozprawy zestawienie publikacji własnych Autora, z podziałem na kategorie. Ułatwia to śledzenie ewolucji zaproponowanej metody. Ilustracje w tekście są liczne, dobrze dobrane do kontekstu i o dobrej jakości graficznej. Mimo swojej znacznej objętości, rozprawa jest zwarta i pozbawiona powtórzeń. Bibliografia obejmuje 215 publikacji, w większości z ostatnich kilku lat. Odwołania w tekście są poprawne. Język rozprawy jest komunikatywny, poprawny stylistycznie i pod względem używanej terminologii.

Niemniej, recenzent zauważył kilka stosunkowo drobnych usterek i błędów:

- w zapisie wzorów (1.7 – 1.8) w rozdziale 1, dla poprawy czytelności, kwantyfikatory powinny występować przed nawiasem obejmującym całe wyrażenie,
- na rys. 5.7, w najniższym lewym bloku jest błędne odwołanie do wzoru (zamiast 5.5 winno być 5.7),
- konstrukcja wzoru (5.7) jest niespójna z treścią komentarza umieszczoną bezpośrednio pod wzorem (*...im mniejsza odległość tym większa waga*?),
- w opisie rys. 5.9 mowa jest o kolumnach, a powinno być o wierszach,

- w algorytmach opisanych w p. 6.1 predykaty konstruowane są bez operowania pojęciem N_4 filtrowanego piksela, a same algorytmy mają wiele wspólnego z metodami morfologii matematycznej obrazów tonalnych, o których nie wspomniano,
- na rys. 6.5 błędnie oznaczono wyjścia z najniższego predykatu w lewej kolumnie,
- na schemacie blokowym z rys.6.20, w obydwóch dolnych predykatach występują odwołania do systemowo numerowanych warunków, jakkolwiek w tekście warunki przedstawione są bez systemowych numerów,
- w kilku miejscach tekstu Autor używa słowa „*liczony*" zamiast słowa „*obliczany*" (słowo „*liczony*" w nieodparty sposób kojarzy mi się z boksem),
- przy tworzeniu map proporcji (wzór 8.1) nie przedyskutowano możliwości wystąpienia bliskich zeru wartości luminancji w rzutowanych obrazach oraz warunku stopu przy iteracyjnym uaktualnianiu map proporcji.

-

4. Konkluzja:

Rozprawa ma zdecydowanie techniczny charakter. Wpisuje się w aktualny nurt prac w perspektywie światowej rozwoju technologii telewizji cyfrowej 3D. Dotyczy ważnych zagadnień praktycznych. Potwierdza duże kompetencje i erudycję Autora.

Tezy rozprawy zostały potwierdzone eksperymentalnie zarówno w aspekcie jakościowej poprawy obrazów syntetyzowanych autorską metodą jak i w aspekcie przydatności proponowanej metody do realizacji systemów wielokamerowych czasu rzeczywistego. Cele pracy, które Autor sformułował na wstępie zostały osiągnięte. Rozprawa stanowi cenny przyczynek do rozwoju technologii telewizji swobodnego dostępu. Wobec powyższego uważam, że opiniowana dysertacja Pana mgr inż. Adriana Dziembowskiego w pełni spełnia wymogi stawiane rozprawom doktorskim przez *Ustawę o stopniach i tytule naukowym*, gdyż przedstawia właściwą koncepcję

rozwiązania istotnego problemu technicznego i jej zaawansowaną (choć wstępną) realizację. Wyniki nawiązują do aktualnego nurtu badań i świadczą o dobrym przygotowaniu kandydata do pracy naukowej. Wnioskuję zatem o dopuszczenie p. mgr inż. Adriana Dziembowskiego do publicznej obrony.

Biorąc pod uwagę ujawnioną w rozprawie i w dokumentacji wyników badań eksperymentalnych, dojrzałość metodologiczną badacza oraz wielki wkład pracy włożony w uzyskanie istotnych wyników wnioskuję jednocześnie o wyróżnienie rozprawy.

(prof. dr hab., inż. Andrzej Kasiński)