

RECENZJA ROZPRAWY DOKTORSKIEJ
DLA
RADY DYSCYPLINY INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA
POLITECHNIKI POZNAŃSKIEJ

Autor rozprawy: **Adam Grzelka**

Tytuł rozprawy doktorskiej: **Systemy wizyjne swobodnego punktu widzenia o wysokiej postrzeganej jakości usług**

Promotor: **prof. dr hab. inż. Marek Domański**

Promotor pomocniczy: **dr hab. inż. Olgierd Stankiewicz**

1. Uwagi wstępne

1.1 Formalna podstawa oceny

Recenzję przygotowano w związku z pismem Dziekana Wydziału Informatyki i Telekomunikacji Politechniki Poznańskiej z dnia 6.08.2024 oraz uchwałą Nr 2024-31-211 Rady Dyscypliny Informatyka Techniczna i i Telekomunikacja z dnia 5.08.2024 r. w sprawie wyznaczenia recenzentów w przewodzie doktorskim mgr inż. Adama Grzelki.

Zadaniem recenzenta jest ocena, czy rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego oraz dowodzi umiejętności samodzielnego prowadzenia pracy naukowej.

1.2 Rozprawa doktorska

Rozprawę doktorską stanowi zwarty manuskrypt liczący 171 stron napisany w języku polskim. Rozprawa zawiera: Streszczenie (0.5 strony), Abstract (0.5 strony), Słownik terminów (1.5 strony), Spis treści (2.5 strony), Rozdział 1 - Wprowadzenie (6 stron), Rozdział 2 – Przegląd stanu wiedzy (17 stron), Rozdział 3 – Metodologia badań (9 stron), Rozdział 4 – Optymalizacja ustawienia kamer w systemach swobodnej nawigacji (21 stron), Rozdział 5 - Wpływ kompresji na jakość wirtualnego widoku (33 stron), Rozdział 6 – Analiza wpływu opóźnień na postrzeganą jakość usługi w wyświetlaczach nagłownych (16 stron), Rozdział 7 – Eksperymentalny system swobodnej nawigacji (20 stron), Rozdział 8 – Podsumowanie (4 strony), Lista publikacji naukowych autora (7 stron) i Bibliografia przedstawiona na 35 stronach.

2. Ocena treści rozprawy i wkładu oryginalnego

2.1. Ocena treści rozprawy

W rozdziale 1 przedstawiono w skondensowanej formie wprowadzenie do zagadnień dotyczących systemów swobodnego punktu widzenia i swobodnej nawigacji czyli systemów umożliwiających użytkownikowi wybór kierunku i miejsca, z którego chce obserwować prezentowaną scenę. Wyświetlane obrazy mogą pochodzić z kamery nagrywającej scenę (kamera rzeczywista) lub też być obrazem (wygenerowanym algorytmicznie) jaki byłby widziany w danym miejscu przez nieistniejącą fizycznie kamerę (kamerę wirtualną). W pracy skupiono się na wykorzystaniu danych rzeczywistych. Systemy tego typu umożliwiają swobodną nawigację w wirtualnej scenie reprezentującej rzeczywiste wydarzenie, dynamicznie zmienne w czasie. Użytkownik do celu swobodnego nawigowania w wirtualnej scenie wykorzystuje terminal użytkownika, służący do prezentacji wirtualnego widoku i sterowania nawigowaną pozycją.

Praca też zagadnienia dotyczące postrzeganej jakości usług (QoE – Quality of Experience), definiowanej w odniesieniu do parametrów subiektywnych (subiektywna jakość obrazu, subiektywna dokuczliwość opóźnień itp.). Skupiono się na pomiarze dwóch typów zniekształceń wpływających na postrzeganą jakość usługi (QoE). Pierwszy związany jest z jakością dostarczanego widoku i reprezentuje wszystkie wady powodujące błędne wartości punktów wirtualnego obrazu. Drugi typ związany jest z kierunkiem i pozycją prezentowanego widoku oraz dokuczliwością i wpływem opóźnień prezentacji treści w systemach swobodnego punktu widzenia.

W rozdziale tym sprecyzowano też cele i tezy rozprawy. Głównym celem pracy jest znalezienie rozwiązań zapewniających postrzeganą wysoką jakość usług w systemach swobodnej nawigacji. W rozdziałach 4, 5, 6 i 7 przeanalizowano, jak poszczególne elementy systemu wpływają na postrzeganą przez użytkownika jakość usługi swobodnej nawigacji.

Sformułowano też 4 tezy rozprawy odniesienie do których znajdzie się w dalszej części recenzji.

Rozdział 2 przedstawia przegląd stanu wiedzy. Przedstawiono schemat systemu swobodnej nawigacji, w którym terminal użytkownika wyświetla wirtualne widoki odpowiadające za usługę swobodnej nawigacji w scenie oraz możliwe jest wykorzystanie różnego typu kamer i urządzeń rejestrujących głębię sceny. W systemie występują:

- serwer reprezentacji wyznaczający reprezentację przestrzenną sceny estymowaną na podstawie widoków z kamer (estymacja głębi),
- serwer brzegowy przesyłający informacje do terminala użytkownika. Może to realizować poprzez syntezę wirtualnego widoku zgodnego z pozycją użytkownika w scenie i wysyłać wyświetlany widok. Może też kodować i wysyłać wszystkie bądź wybrane obrazy bez przetwarzania,
- terminal użytkownika wyświetlający obraz na ekranie lub umożliwiający skorzystanie z urządzenia pozwalającego na uzyskanie efektu trójwymiarowego. Może też lokalnie realizować syntezę wirtualnego widoku.

W rozdziale tym zasygnalizowano też:

- prace dotyczące kompresji pojedynczego widoku jak i wizji wielowidokowej, trójwymiarowej oraz wszechkierunkowej,
- prace związana z estymacją map głębi,
- syntezą widoków wirtualnych,
- rozmieszczeniem kamer w scenie,
- synchronizacją.

Rozdział ten nie wnosi do przedstawianych w rozprawie zagadnień nowych treści jednak może być dobrym wprowadzeniem do zrozumienia zagadnień związanych z systemami swobodnej nawigacji.

Rozdział 3 przedstawia zbiór sekwencji testowych wykorzystywanych podczas pracy nad rozprawą doktorską oraz omawia metryki PSNR i IV-PSNR. Rozważania przedstawione w tym rozdziale są znane i nie wnoszą żadnych istotnych rezultatów do rozprawy.

Rozdział 4 poświęcony jest znalezieniu zależności określającej jak położenie kamer w scenie wpływa na jakość systemu swobodnej nawigacji. Rozpatrzono estymację dla pary kamer, a następnie rozszerzono analizę na przypadek ogólny. Estymacja map głębi jest możliwa dla wszystkich obszarów sceny, jeśli zostały one zarejestrowane przez co najmniej dwie kamery systemu. Próbki, które nie są widoczne w co najmniej dwóch obrazach, będą nazywane przysłoniętymi. Dla takich obszarów nie jest możliwe wyznaczenie mapy głębi poprzez znalezienie odpowiadających próbek w widokach pozostałych kamer. Wartości mapy głębi dla obszarów przysłoniętych mogą zostać wyznaczone na podstawie obszarów sąsiednich, co powoduje spadek jakości wyznaczanych danych. Modyfikacji poddawana jest pozycja kamer środkowych tak aby skrajne kamery systemu rejestrowały ten sam obszar, który stanowi całą scenę. Celem optymalizacji ustawienia kamer będzie uzyskanie wysokiej jakości widoków wirtualnych na podstawie kompromisowej odległości pomiędzy kamerami. Do tego celu użyto metrykę PSNR. Wyniki dla metryki IV-PSNR są podobne.

W rozdziale rozpatrywano model systemu wielokamerowego i analizowano wpływ odległości pomiędzy kamerami na jakość wirtualnego widoku. Wprowadzono model średniej jakości syntezowanego widoku wykorzystywany do oszacowania wpływu przysłonieć w scenie i odległości pomiędzy kamerami na jakość wirtualnego widoku. Autor przedstawia wyniki dotyczące wpływu odległości pomiędzy kamerami na jakość estymowanej mapy głębi i wirtualnych widoków. Mapa głębi może być estymowana z dużą dokładnością, gdy odległość pomiędzy kamerami jest duża. Dlatego też, ze względu na różnice głębi, w przypadku wielu kamer, głębia powinna być estymowana z wykorzystaniem dwóch kamer, które są najdalej położone w systemie.

W scenach z przysłonieniami poszczególne obszary mogą nie być rejestrowane przez wszystkie kamery a do estymacji mapy głębi należy wybrać te najbardziej oddalone od siebie czyli skrajne.

Wyniki teoretyczne uzyskane przez Autora pokazują, że stosowanie par kamer zamiast równomiernego ich rozmieszczenia jest korzystne, gdy procentowy udział przysłoniętych powierzchni jest większy niż około 25% przy czym ustawienie kamer w parę zmniejsza dystans pomiędzy nimi o 60%.

Autor pokusił się też o zweryfikowanie zależności teoretycznych pomiędzy jakością wirtualnego obrazu a rozmieszczeniem kamer (stałe położenie kamer zewnętrznych i zmienne położenie kamer wewnętrznych).

Wynikiem prac badawczych zaprezentowanych w tym rozdziale dotyczących wpływu odległości pomiędzy kamerami i przysłonieć w scenie na jakość wirtualnego widoku są następujące konkluzje:

- liczba przysłonieć w scenie oraz zmiana odległości pomiędzy kamerami wpływa na średnią jakość syntezowanych widoków (wg. mojej opinii jest to konkluzja oczywista),
- umieszczanie kamer w parach daje zysk jakości w przypadku, kiedy w scenie występuje wiele przysłonieć. Wykazano, że pary kamer należy stosować, kiedy procentowa liczba przysłoniętych próbek obrazu w syntezowanych widokach dla kamer rozstawionych równomiernie przekroczy około 25%.

Rozdział 5 opisuje problem określenia wpływu stratnej kompresji widoków realizowanej przed procesem estymacji map głębi na jakość swobodnej nawigacji. Autor za cel badań przyjął udowodnienie, że sekwencja wielowidokowa może być skompresowana przed procesem estymacji map głębi, a także zbadanie wpływu tej kompresji na jakość wirtualnego widoku.

Autor przeprowadził szereg eksperymentów dla różnych wariantów kodowania:

- eksperyment wstępny przeprowadzony na grupie wielowidokowych sekwencji dla kompresji wybranych widoków koderami pojedynczego widoku, takimi jak MPEG-2, AVC oraz HEVC. Wybrane źródłowe widoki są kodowane oddzielnie i używane jako dane wejściowe dla oprogramowania estymacji głębi oraz syntezy wirtualnych widoków,
- eksperyment zasadniczy wykorzystujący techniki kompresji takie jak MPEG Immersive Video (MIV) oraz nową metodę oceny jakości widoków zszyntezowanych za pomocą metryki IV-PSNR.

Przeprowadzony eksperyment wstępny wykazał, że zastosowanie kompresji z małą wartością współczynnika parametru kwantyzacji nie obniża znacząco jakości syntezowanych wirtualnych widoków. Dla wszystkich testowanych koderów (MPEG2, AVC i HEVC) różnica między jakością wirtualnego widoku syntezowanego przy użyciu nieskompresowanej i skompresowanej wizji nie była większa niż 1 dB. Oznacza to, że sekwencja wielowidokowa może zostać skompresowana co zapewni kilkusetkrotne zmniejszenie wymaganej prędkości bitowej oraz zmianę jakości wirtualnego widoku nie większą niż 0,5 dB przy założeniu, że jakość przesyłanych widoków ma mały współczynnik kwantyzacji. W eksperymencie MIV wykorzystano sekwencje dookólne, perspektywiczne, treści generowane komputerowo oraz sekwencje naturalne zarejestrowane przez rzeczywiste systemy wielokamerowe. Zaprezentowane wyniki wykazują, że zarówno w przypadku wykorzystania techniki ISO/IEC, dotyczącej kompresji wizji wszechogarniającej (MIV), jak i w ogólnym przypadku jakość wirtualnych widoków zależy od jakości widoku przesłanego. Zastosowana technika kompresji w niewielkim stopniu wpływa na jakość wirtualnych widoków, a tym samym na postrzeganą jakość usługi przez odbiorcę systemu.

Rozdział 6 omawia wpływ opóźnień transmisji na postrzeganą jakość usługi w wyświetlaczach nagłownych. Prezentowany widok powinien zmieniać się zgodnie z rzeczywistą pozycją i kierunkiem głowy użytkownika czyli opóźnienie reakcji pomiędzy zmianą rzeczywistej pozycji i/lub kierunku obrotu głowy użytkownika a wyświetleniem zaktualizowanego widoku wirtualnego T powinno być możliwie małe. Dwa sposoby dostarczania danych do terminala użytkownika to:

- dostarczanie pełnej reprezentacji sceny niezależnej od pozycji i kierunku patrzenia,
- transmisja informacji o pozycji użytkownika i kierunku patrzenia do serwera brzegowego oraz przygotowaniu wirtualnego, dedykowanego widoku. Możliwa jest optymalizacja składników opóźnienia T . Główne składniki tego opóźnienia to zazwyczaj czas transmisji danych przez sieć będące czasem transmisji informacji o pozycji i kierunku patrzenia oraz czasem transmisji wirtualnego widoku. Autor rozpatrywał 3 scenariusze w których występowało opóźnienie transmisji pozycji i kierunku patrzenia użytkownika, opóźnienie transmisji pozycji użytkownika i opóźnienie transmisji pozycji i kierunku nie występowały. Eksperyment przeprowadzono w formie testów subiektywnych dla dwóch scenariuszy, w których występują opóźnienia pozycji i/lub kierunku patrzenia użytkownika. Są to rozwiązania zapewniające transmisje dedykowanych danych do terminala oraz niewymagające pełnej syntezy.

Autor zaimplementował symulator opóźnienia pozycji i kierunku patrzenia użytkownika w scenie wykorzystywany podczas testów subiektywnych.

Rozdział 7 prezentuje eksperymentalny system swobodnej nawigacji składający się wielokamerowego systemu akwizycji oraz serwera swobodnej nawigacji. System wielokamerowy jest systemem, pozwalającym na rejestrowanie danych rzeczywistych (np. estymacji map głębi) wykorzystywanych przez serwer swobodnej nawigacji. Zapewnia on usługę swobodnego punktu widzenia, czyli przygotowuje wirtualne widoki, które są przesyłane przez sieć do terminali użytkowników.

Omawiana implementacja może być bazą do dalszych prac badawczych szczególnie w zakresie przeniesienia części obliczeń na procesor graficzny (GPU, ang. Graphics Processing Unit), zastosowania wydajniejszych technik kodowania oraz protokołów transmisji dostosowanych do usług realizowanych w czasie rzeczywistym. Możliwa jest implementacja serwera z wykorzystaniem wydajnego komputera klasy PC.

Rozdział 8 prezentuje wnioski o charakterze teoretycznym, eksperymentalnym, jak i konstrukcyjnym dotyczące monografii doktorskiej w szczególności postrzeganej jakości w systemach swobodnej

nawigacji. Rozważania teoretyczne dotyczyły opracowania modelu wpływu przysłonieć na jakość wirtualnego widoku (rozdział 4) oraz opracowania zależności jakości widoku wirtualnego w funkcji jakości widoków przysłanych (rozdział 5). W rozdziałach 4, 5 oraz 6 opisano czasochłonne eksperymenty przeprowadzone przez autora z wykorzystaniem zbioru sekwencji testowych. Pozwoliły one na walidację modeli teoretycznych przedstawionych w rozdziale 4. Prace eksperymentalne przedstawione w rozdziale 5 były podstawą do zaproponowania modelu wpływu jakości przesyłanego widoku na jakość wirtualnego widoku. Przeprowadzono również testy subiektywne z wykorzystaniem wyświetlacza nagłownego w celu zbadania wpływu opóźnień w systemie swobodnego punktu widzenia. Eksperyment ten został przedstawiony w rozdziale 6. W rozdziale 7 pracy przedstawiono nowy, kompletny system wielokamerowy oraz serwer swobodnej nawigacji, które stanowią osiągnięcie konstrukcyjne.

2.2. Ocena oryginalności rozprawy.

W rozprawie udowodniono zaprezentowane w rozdziale 1 tezy rozprawy. Tezę 1 udowodniono w rozważaniach rozdziału 4. Tezę 2 rozważano w rozdziale 5. Tezę 3 omówiono w rozdziale 6 natomiast tezę 4 w rozdziale 7.

Za oryginalne aspekty rozprawy uważam:

- podjęcie tematyki postrzeganej jakości w systemach swobodnej nawigacji. Zagadnienia te mają zarówno charakter teoretyczny, eksperymentalny, jak i konstrukcyjny. Autor opisał teoretyczne aspekty dotyczące opracowania modelu wpływu przysłonieć na jakość wirtualnego widoku (rozdział 4) oraz opracowania zależności jakości widoku wirtualnego w funkcji jakości widoków przysłanych (rozdział 5). W rozdziałach 4, 5 oraz 6 opisano eksperymenty przeprowadzone przez Autora z wykorzystaniem zbioru sekwencji testowych. Pozwoliły one na walidację modeli teoretycznych przedstawionych w rozdziale 4. Prace eksperymentalne przedstawione w rozdziale 5 były podstawą do zaproponowania modelu wpływu jakości przesyłanego widoku na jakość wirtualnego widoku. Ponadto, przeprowadzono testy subiektywne z wykorzystaniem wyświetlacza nagłownego w celu zbadania wpływu opóźnień w systemie swobodnego punktu widzenia. Eksperyment ten został przedstawiony w rozdziale 6. W rozdziale 7 pracy przedstawiono nowy, kompletny system wielokamerowy oraz serwer swobodnej nawigacji, które stanowią osiągnięcie konstrukcyjne.
- opracowanie modelu teoretycznego wpływu przysłonieć w scenie na jakość wirtualnego widoku.
- potwierdzenie poprawności dwóch teoretycznych modeli wpływu przysłonieć w scenie i odległości pomiędzy kamerami na jakość wirtualnego widoku.
- potwierdzenie, że kompresja widoków przed estymacją map głębi nie wpływa znacząco na jakość wirtualnego widoku, co pozwala na jej stosowanie w systemach swobodnego punktu widzenia.
- stwierdzenie, że jakość wirtualnego widoku zależy głównie od jakości przesyłanych widoków i w stosunkowo niewielkim stopniu od zastosowanej techniki kompresji.

W tym aspekcie uważam, że Autor rozprawy podjął ważny, aktualny a zarazem trudny problem naukowy. Zaproponował też pewne autorskie rozwiązania oraz przedstawił oryginalne przyczynki pozwalające na kontynuację badań w przedstawionym w rozprawie kierunku.

3. Poprawność redakcyjna rozprawy i uwagi krytyczne

Redakcja pracy jest poprawna. Drobne uwagi dotyczą:

- braku jawnie sformułowanych i deklarowanych oryginalnych przyczynków,
- stosowania w pewnych fragmentach pracy oznaczeń z innych publikacji.

Rozważania przedstawione w rozdziale 3 są znane i nie wnoszą żadnych istotnych rezultatów do rozprawy.

Opisane powyżej uwagi nie mają istotnego wpływu na wagę przedstawionych rozwiązań i nie obniżają wartości pracy.

4. Ocena końcowa

Autor wykazał się dużą wiedzą w zakresie tematyki rozprawy, umiejętnością pracy naukowej oraz znajomością metod badawczych. Wyniki pracy świadczą o dobrym przygotowaniu Autora do pracy naukowej.

Tematyka pracy lokuje się w szerokim nurcie prac zespołu kierowanego przez promotora prof. dr hab. inż. Marka Domańskiego. Prace te w wielu aspektach naukowych częściowo się pokrywają, jednak każda zawiera elementy oryginalne.

Biorąc pod uwagę wszystkie elementy recenzowanej pracy uważam, że opiniowana dysertacja zasługuje na wyróżnienie.

Stwierdzam, że recenzowana rozprawa doktorska Pana Adama Grzelki spełnia warunki określone w art. 13.1 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. nr 65 poz. 595 z późn. zmianami) i wnioskuję do Rady Dyscypliny Naukowej Informatyki Technicznej I Telekomunikacji Politechniki Poznańskiej o dopuszczenie Pana Adama Grzelki do dalszych etapów przewodu doktorskiego.

11.10.2024
data sporządzenia recenzji


podpis recenzenta