

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Wydział Informatyki

STRESZCZENIE ROZPRAWY DOKTORSKIEJ

METODY MODELOWANIA PERCEPCJI GŁĘBI ORAZ WIDZENIA KIERUNKOWEGO NA POTRZEBY OBRAZOWANIA KOMPUTEROWEGO

Bartosz Bazyluk

Promotor:

dr hab. inż. Radosław Mantiuk, prof. ZUT

Recenzenci:

prof. dr hab. inż. Przemysław Rokita
Politechnika Warszawska

dr hab. inż. Adam Wojciechowski
Politechnika Łódzka

Szczecin 2017

Spis treści

1	Aktualność problemu	2
2	Przedmiot badań	3
3	Metody naukowe	4
4	Główny cel rozprawy	5
5	Zadania do rozwiązania	5
6	Wartość teoretyczna	6
7	Wartość praktyczna	6
8	Akceptacja przez społeczność naukową	7
9	Na obronę wynosi się	8
10	Struktura i układ pracy	8
11	Zawartość pracy	9
11.1	Modelowanie kierunkowości widzenia	9
11.1.1	Problem jakości danych	10
11.1.2	Metoda GDOT	12
11.1.3	Zastosowania metody GDOT	19
11.2	Interaktywna symulacja głębi ostrości	25
11.2.1	Implementacja i ewaluacja rezultatów	26
11.2.2	Łagodzenie konfliktu konwergencji-akomodacji	29
11.3	Głębia w obrazach o poszerzonym zakresie dynamiki luminancji	31
11.3.1	Badanie zależności kontrastu i wrażenia głębi	31
11.3.2	Kontrast i głębia a poczucie realizmu	31
11.4	Podsumowanie	34
Bibliografia		35
	Publikacje własne	35
	Wybrane pozycje literaturowe	37

1 Aktualność problemu

Komputerowa synteza grafiki jest zagadnieniem, które zawsze cechowało dążenie do realizmu otrzymywanych rezultatów. Należy zauważyć, że zdefiniowanie pojęcia realizmu obrazu uzyskanego w wyniku procesu komputerowego zdaje się nie być rzeczą trywialną. W ujęciu zaproponowanym przez Ferwerdę, mianem tym określa się dążenie do zachowania cech zbliżonych do rzeczywistego świata w trzech rozróżnialnych standardach: fizykalnym, fotorealistycznym i funkcjonalnym [Fer03]. Ujęcie fotorealistyczne dotyczy tu prób uzyskania obrazu, który wywoływałby taką samą reakcję systemu wizyjnego u obserwatora, jak oglądanie rzeczywistej sceny. Wieloaspektowość tego zagadnienia obejmuje szerokie spektrum elementów związanych z ludzką percepcją wzrokową: postrzegania koloru, kontrastu, układu przestrzennego czy głębi.

Należy jednak pamiętać, że nawet najdoskonalej zsyntezowana komputerowo, statyczna wizualizacja sceny, nie jest w stanie spełnić kryterium fotorealizmu gdy wyjście systemu komputerowego posiada ograniczoną zdolność do reprodukcji cech obrazu. Niniejsza praca koncentruje się na próbie dostarczenia ludzkiemu systemowi wizyjnemu takich bodźców, które wywołają wrażenie głębi, a więc i świadomości przestrzennej, zbliżone do obserwacji sceny w świecie rzeczywistym. W tym celu wykorzystywane są narzędzia takie, jak wyświetlanie stereoskopowe, wyświetlacze o poszerzonym zakresie dynamiki luminancji oraz systemy okulograficzne pozwalające na dynamiczne dostosowanie cech obrazu do sposobu jego oglądania.

W celu zamodelowania związków pomiędzy elementami systemu percepcji a wrażeniem głębi oglądanego obrazu konieczne było wykonanie badań o charakterze podstawowym. Eksperymentalnie określono relację pomiędzy kontrastem, a postrzeganiem odległości obiektów w przypadku wyświetlacza o poszerzonym zakresie dynamiki luminancji oraz potwierdzono znane wcześniej zjawisko zaobserwowane w przypadku wyświetlaczy o wąskim zakresie dynamiki. Dokonano określenia związku pomiędzy występowaniem obrazowania zjawiska głębi ostrości, a wrażeniem głębi. Opisano także wpływ kontrastu dla obrazów stereoskopowych w warunkach poszerzonego zakresu dynamiki luminancji na poczucie realności obserwowanego obrazu.

Implementacyjna część pracy dotyczyła opracowania sposobu realizacji interaktywnej symulacji zjawiska głębi ostrości. W oparciu o pochodzącą z okulografu informację o elemencie trójwymiarowej sceny na którym skupiony jest wzrok obserwatora, symulowany jest proces adaptacji soczewki oka do odległości na jakiej obiekt ten się znajduje. Badano też możliwość wpływu proponowanej metody na złagodzenie konfliktu konwergencji i akomodacji w przypadku obrazów stereoskopowych.

Wykorzystanie technik okulograficznych w czasie rzeczywistym do zmiany cech prezentowanego obrazu jest podejściem nowatorskim, opisanym w literaturze w bardzo ograniczonym zakresie. Koniecznością okazało się więc przezwyciężenie istotnych problemów związanych z dokładnością identyfikacji elementów obrazu na których spoczywa uwaga obserwatora w przypadku dynamicznej wizualizacji interaktywnych środowisk rzeczywistości wirtualnej. Zakres aplikacji opracowanej metody jest bardzo szeroki i wykracza poza główny nurt pracy.

2 Przedmiot badań

Obserwator oglądający obraz będący wizualizacją pewnej sceny, dokonuje jego oceny kierując się licznymi kryteriami. Kryteria te są często nieświadome i bazują na doświadczeniu związanym z codziennym *postrzeganiem* (ang. *perceiving*) otaczającego świata. Doświadczenie to budowane jest na podstawie wrażeń i reakcji, które zazwyczaj wywoływane są przez *bodźce wzrokowe* (ang. *visual stimuli*) w postaci przestrzennych obiektów z realnego otoczenia. Kluczowym elementem, na który należy tu zwrócić uwagę, jest właśnie *poczucie przestrzenności* środowiska (ang. *spatial perception*): zdolność do świadomego rozróżniania tworzących je brył oraz estymacji ich odległości zarówno w ujęciu bezwzględnym, jak i względnym. Zdolność, która pozwala na codzienną interakcję z przedmiotami i świadome funkcjonowanie człowieka w jego otoczeniu.

Świat rzeczywisty charakteryzuje się znaczącym zróżnicowaniem odległości od obserwatora, w zakresie których mieszczą się tworzące go obiekty. Naturalnym więc jest, że cechujące każdego widza subiektywne *poczucie realności* otoczenia (ang. *realism perception*), wymaga od obrazu naśladującego rzeczywisty świat, by ten tworzył w świadomości trójwymiarową reprezentację wizualizowanej sceny. By to umożliwić w przypadku obrazów generowanych komputerowo, stosowane są różnorodne sposoby ich syntezy i prezentacji, uwzględniające *głębnię* (ang. *depth*) przedstawianego wycinka wirtualnego świata.

Oddziaływanie na ludzki system postrzegania w sposób, który ma na celu wywołać wrażenia jak najbardziej zbliżone do tych doświadczanych w trakcie obcowania ze światem rzeczywistym, jest zadaniem niezwykle złożonym i problematycznym. Wiąże się to z trudnościami technicznymi dotyczącymi zarówno metod syntezy obrazów, jak i metod ich wyświetlania.

Człowiek buduje swoją umysłową reprezentację przestrzennego układu sceny na podstawie zawartych w oglądanym obrazie *wskazówek głębi* (ang. *depth cues*). Tym pojęciem określa się elementy bodźców tworzące lub wzmacniające wrażenie różnic głębokości pomiędzy poszczególnymi składnikami obrazu, takie jak stereoskopia (wskazówka binokularna), czy perspektywa, akomodacja, kontrast, przesłonięcie i ruch obiektów (wskazówki monokularne) [ZJM09]. Celem działania wyidealizowanych komputerowych systemów syntezy i prezentacji obrazu byłoby więc dostarczenie ludzkiemu aparatowi widzenia takiej stymulacji, która będzie zawierała w sobie kompletny zestaw wskazówek znanych z realnego otoczenia.

Współczesne sposoby wizualizowania wirtualnych scen komunikują wskazówki głębi jedynie w ograniczonym stopniu. Metody dostarczania obserwatorowi zróżnicowanych wrażeń obuoczných są niedoskonałe. Wyświetlanie stereoskopowe wiąże się z redukcją kontrastu, obarczone jest zakłóceniami dostarczanej informacji i stawiając wymuszoną konwergencję oczu w konflikcie z akomodacją do płaszczyzny ekranu, prowadzi do nienaturalnego zachowania i w konsekwencji zmęczenia aparatu widzenia. Z kolei samo zjawisko akomodacji i powiązanej z nią głębi ostrości jest często pomijane lub ukazywane w sposób naśladujący wyreżyserowany film, zamiast naśladowania rzeczywistego, reagującego na kierunek spojrzenia działania cechującego system widzenia człowieka. Natomiast niedostateczny zakres dynamiki urządzeń wyświetlających obraz skutkuje niewystarczającymi poziomami jasności i zbyt małym

kontrastem obrazu.

Ograniczenia te stanowią przeszkodę w dostarczeniu użytkownikowi wrażeń mogących skutkować u niego poczuciem pełnego uczestnictwa w wirtualnej rzeczywistości. Nie pozwalają one na zatarcie granicy pomiędzy syntetycznymi bodźcami, a odczuwaniem obrazowanego za ich pomocą środowiska. Dążąc do uzyskania jak najbardziej przekonującej symulacji rzeczywistości konieczne staje się więc poszukiwanie sposobów na to, by próbować przezwyciężyć te ograniczenia oraz wykorzystać dostępne środki tak, aby jak najdoskonalej przekazać niezbędną informację o przestrzenności sceny.

3 Metody naukowe

Praca ma charakter eksperymentalno-analityczny. Z uwagi na przedmiot badań, podstawową metodą badawczą jest prowadzenie eksperymentów percepcyjnych dotyczących ludzkiego sposobu postrzegania obrazów syntezowanych i wyświetlanych przez graficzne systemy komputerowe. Wyniki eksperymentów są poddawane analizie, na podstawie której powstają modele umożliwiające późniejsze wykorzystanie badanych zależności. Te zaś weryfikowane są pod kątem ich poprawności.

Przykładem zastosowania tego podejścia jest zamodelowanie dokonywanej przez obserwatorów oceny jakości identyfikacji obiektu, na którym skupiona jest ich uwaga wzrokowa. Na podstawie wyników eksperymentu percepcyjnego powstała metryka jakości, która stała się podstawą dla zaprojektowania i kalibracji metody filtracji danych okulograficznych. Metoda ta jest następnie wykorzystywana w licznych aplikacjach (symulacja głębi ostrości, analiza uwagi wizualnej w materiałach wideo, elementy interfejsu użytkownika gry komputerowej), których skuteczność działania jest eksperymentalnie weryfikowana. Inne zadania, w których użyto podobnej metodologii, to modelowanie zależności głębi od kontrastu wyświetlacza stereoskopowego oraz pomiar dokładności urządzeń śledzących kierunek spojrzenia.

Operując w obszarze subiektywnych odczuć użytkownika systemu komputerowego, ewaluacja proponowanych rozwiązań opiera się też na ankietyzacji badanych osób i analizie statystycznej uzyskanych wyników. W ten sposób rozpatrywany jest wpływ opisanych w pracy technik na poczucie zanurzenia w wirtualnym środowisku, a także określane są preferencje obserwatorów odnośnie zaimplementowanych rozwiązań.

W pracy badany jest ludzki układ postrzegania, który nosi znamiona systemu czasu rzeczywistego o ciągłej charakterystyce pracy. Synteza i wyświetlanie obrazów naśladujących rzeczywistość, niosących wiarygodną informację o głębi sceny, stanowi wyzwanie z uwagi na szybkość działania tego układu. W ramach pracy konieczne było rozwiązanie szeregu związanych z tym problemów technicznych. W toku rozprawy dokonano więc implementacji wielu opracowanych rozwiązań, takich jak działająca w czasie zbliżonym do rzeczywistego metoda identyfikacji elementu sceny będącego przedmiotem uwagi obserwatora, bazująca na danych okulograficznych metoda symulacji głębi ostrości w oparciu o podejście *post-processing*, czy też metoda analizy uwagi obserwatora oglądającego obraz wideo, wspomagana techniką śledzenia przepływu optycznego.

Przeprowadzone eksperymenty

W ramach rozprawy przeprowadzono następujące badania eksperymentalne:

- Wpływ kontrastu obiektów na postrzeganie względnej różnicy odległości w obrazach wyświetlanych na ekranie o wąskim zakresie dynamiki luminancji (*LDR*).
- Wpływ kontrastu obiektów na postrzeganie względnej różnicy odległości w obrazach wyświetlanych na ekranie o poszerzonym zakresie dynamiki luminancji (*HDR*).
- Wpływ kontrastu na wrażenie realności stereoskopowego obrazu w przypadku możliwości porównania z rzeczywistą referencją.
- Wpływ wizualizacji zjawiska głębi ostrości na postrzeganie względnej różnicy odległości w obrazach stereoskopowych.
- Wpływ interaktywnej wizualizacji zjawiska głębi ostrości w środowisku wirtualnej rzeczywistości na poczucie zanurzenia u obserwatora.

4 Główny cel rozprawy

Na potrzeby rozważań ujętych w rozprawie postawiono następującą tezę:

Wykorzystanie modeli mechanizmu percepcji głębi człowieka umożliwia poprawę jakości wizualizacji obrazów trójwymiarowych. Dotyczy to takich mechanizmów jak: kierunkowość widzenia, czułość na kontrast w szerokim zakresie dynamiki luminancji, mechanizm akomodacji, widzenie stereoskopowe oraz subiektywne wrażenie realizmu obrazu.

Za główny cel pracy postawiono zbadanie wybranych aspektów sposobu widzenia człowieka w kontekście ich użyteczności dla komputerowych metod wizualizacji głębi obrazu, a następnie zaproponowanie i sprawdzenie skuteczności rozwiązań umożliwiających wykorzystanie pozyskanej wiedzy w procesie syntezy oraz prezentacji grafiki komputerowej.

Do rozpatrywanych aspektów widzenia należą przede wszystkim: powiązana z akomodacją głębia ostrości, widzenie stereoskopowe oraz postrzeganie kontrastu. Dodatkowo istotnym narzędziem staje się estymowanie kierunku spojrzenia w czasie zbliżonym do rzeczywistego, które umożliwia realizację technik wzbogacających wyświetlanie obrazu.

5 Zadania do rozwiązania

W trakcie prac nad rozprawą konieczne stało się rozwiązanie następujących zadań, z których każde jest złożonym problemem wymagającym wielopłaszczyznowego rozpatrzenia:

- Zbadanie skuteczności metod filtracji danych okulograficznych na potrzeby identyfikacji elementu, na którym spoczywa uwaga obserwatora w środowiskach trójwymiarowych i zaproponowanie metody dedykowanej dynamicznym wizualizacjom wirtualnych środowisk.
- Zbadanie możliwych zastosowań wiedzy o punkcie skupienia wzroku w wizualizacji wirtualnego środowiska zarówno w kontekście możliwości przyspieszenia renderowania, jak i elementów interfejsu użytkownika.
- Opracowanie metody interaktywnej, opartej na danych okulograficznych symulacji zjawiska głębi ostrości, zbadanie jej wpływu na odbiór wizualizacji wirtualnego środowiska przez użytkownika oraz zbadanie możliwości zredukowania za jej pomocą dyskomfortu konwergencji i akomodacji w wyświetlaczach stereoskopowych.
- Zbadanie wpływu kontrastu w stereoskopowych wyświetlaczach o poszerzonym zakresie dynamiki luminancji na postrzeganie głębi oraz wrażenie realności oglądanego obrazu.

6 Wartość teoretyczna

Przeprowadzone badania w zakresie odpowiedzi ludzkiego systemu percepcji wizyjnej miały charakter badań podstawowych. Można wyróżnić następujące zagadnienia, które zostały potwierdzone, opisane i zamodelowane podczas powstawania tejże pracy:

- Potwierdzono zależność różnicy wrażenia głębi od względnej różnicy kontrastów obiektów w warunkach małej dynamiki zakresu luminancji.
- Zbadano, rozpoznano i zamodelowano zależności związane z wpływem różnicy kontrastu na wrażenie głębi w obrazach o poszerzonym zakresie dynamiki luminancji.
- Podobnie zbadano zależność pomiędzy występowaniem zjawiska głębi ostrości, a postrzeganiem relatywnej odległości dwu obiektów.
- Wykazano, że u niektórych obserwatorów nadmierny kontrast wyświetlanego obrazu stereoskopowego powoduje nasilenie wrażenia jego realności.
- Określono związek pomiędzy interaktywnym charakterem symulacji zjawiska głębi ostrości w komputerowo syntezowanym obrazie wirtualnej rzeczywistości, a poczuciem zanurzenia u obserwatora.

7 Wartość praktyczna

Implementacyjny charakter wielu z poruszanych aspektów pozwolił na praktyczne zastosowanie wyników badań. Do szczególnie istotnych osiągnięć praktycznych zalicza się:

- Utworzenie wydajnej implementacji algorytmu filtracji danych eyetrackingowych w czasie rzeczywistym o nazwie *GDOT* (ang. *Gaze-Driven Object Tracking*).
- Implementację opartą o programowalne układy graficzne metody interaktywnej symulacji zjawiska głębi ostrości współdziałającej w czasie rzeczywistym z okulografem.
- Budowę rozszerzeń interfejsów użytkownika wykorzystujących wiedzę o uwadze wizualnej na potrzeby interakcji z wirtualną rzeczywistością.
- Wykorzystanie wiedzy o charakterze uwagi wizualnej podczas oglądania dynamicznego obrazu, do opracowania efektywnej metody identyfikacji obiektów przyciągających uwagę na przestrzeni czasu.

8 Akceptacja przez społeczność naukową

Merytoryczny wkład pracy w przedmiotową dziedzinę opiera się na publikacjach autora. Pojawiły się one na łamach następujących wydawnictw:

- Computer Graphics Forum [MBM13]
- Lecture Notes in Computer Science (*trzykrotnie*) [MBT11, MKNB12, BM14a]
- Proceedings of ACM Symposium on Applied Perception (*dwukrotnie*) [Pet14, Duc14]
- Proceedings of Eurographics [MCC⁺11]
- Pomiar Automatyka Kontrola [TB10]
- Proceedings of Central European Seminar on Computer Graphics [Baz10]

Część wyników prac została także opublikowana w postaci dwóch rozdziałów monografii *Widziane inaczej. Z polskich badań eyetrackingowych* [BM14b, MB14], jednej z bardzo nielicznych polskojęzycznych publikacji dotyczących z okulografii.

Przeprowadzane badania były dyskutowane podczas wystąpień plenarnych na następujących konferencjach naukowych:

- Eurographics 2013 (*EG*)
- International Conference on Computer Vision and Graphics 2014 (*ICCVG*)
- Polska Konferencja Eyetrackingowa 2015 i 2016 (*PKET*)
- Serious Games Development and Applications 2011 (*SGDA*)
- Advanced Computer Systems 2012 (*ACS*)
- Central European Seminar on Computer Graphics 2010 (*CESCG*)

- Reprogramowalne Układy Cyfrowe 2010 (*RUC*)
- Sejmik Młodych Informatyków 2012 (*SMI*)

Ponadto metoda filtracji danych okulograficznych powstała podczas badań związanych z tematyką pracy stała się przedmiotem dwóch zgłoszeń patentowych: *Sposób identyfikacji fiksacji wzroku użytkownika* (Polska, *EPO*) oraz *System and method for probabilistic object tracking over time* (Stany Zjednoczone, *IPC PCT*).

9 Na obronę wynosi się

Osiągnięcia deklarowane do obrony:

- Zaprojektowanie metody interaktywnej symulacji zjawiska głębi ostrości w aplikacjach wirtualnej rzeczywistości, działającej w oparciu o dane okulograficzne.
- Implementacja tej metody z zastosowaniem technik grafiki komputerowej czasu rzeczywistego i możliwości programowalnych procesorów graficznych.
- Zaproponowanie i implementacja elementów rozszerzających interfejs użytkownika gry komputerowej opartych o dane okulograficzne.
- Wykazanie wpływu kontrastu na postrzeganie głębi w obrazach o poszerzonym zakresie dynamiki luminancji.
- Implementacja metody identyfikacji obiektu będącego przedmiotem uwagi wizualnej obserwatora w wizualizacji dynamicznej sceny trójwymiarowej.
- Wykorzystanie tej metody na potrzeby analizy przebiegu uwagi wizualnej obserwatora podczas oglądania obrazu wideo.

10 Struktura i układ pracy

Na rozprawę składają się trzy główne rozdziały.

Rozdział 1 zawiera opis podstaw dotyczących percepcji człowieka i postrzegania przez niego przestrzeni na podstawie obrazów, które są mu przedstawiane przez graficzny system wirtualnej rzeczywistości. Rozważania dotyczą koncepcji realizmu symulowanego świata i jego pocucia u użytkownika systemu, jak i jej powiązań z postrzeganiem przestrzenności sceny. Wskazana zostaje istotność aspektu kierunkowości widzenia obserwatora, jako jednego z mechanizmów uwagi wizualnej, który przekłada się na zdolność do odbioru informacji przestrzennej o oglądanej scenie. Zawarte w tym rozdziale treści stanowią tło dla właściwej pracy badawczej, opisanej w dalszej części rozprawy.

Rozdział 2 jest w całości poświęcony problematyce kierunkowości widzenia, rozumianej jako narzędzie pozwalające wprowadzić nowe elementy do symulacji wirtualnej rzeczywistości. Rozdział dotyczy technicznych możliwości śledzenia kierunku

spojrzenia za pomocą okulografii. Przedstawione zostają współczesne problemy tego obszaru oraz, opracowana w ramach rozprawy, innowacyjna technika o nazwie *GDOT*, która znacząco poprawia jakość identyfikacji tych wyświetlanych obiektów, na które spogląda w danej chwili użytkownik. Proponowane są zastosowania tej techniki, wykraczające także poza obszar wirtualnej rzeczywistości. Rozwiązania te zostają poddane ewaluacji.

Natomiast Rozdział 3 dotyczy sposobów wpływania na występujące u użytkownika poczucie przestrzenności za pomocą technik syntezy i prezentacji obrazu. Proponowana jest idea oraz implementacja interaktywnej symulacji zjawiska głębi ostrości, która wykorzystuje zbierane na bieżąco dane okulograficzne. Badana jest skuteczność tego rozwiązania. Ponadto badana jest próba wykorzystania go w celu złagodzenia niedogodności występujących podczas oglądania obrazów stereoskopowych.

Dalsze badania opisane w Rozdziale 3 dotyczą stereoskopowych wyświetlaczy o poszerzonym zakresie dynamiki luminancji. Modelowana jest zależność pomiędzy kontrastem a postrzeganiem różnicy odległości. Kontrast i obrazowanie stereoskopowe są także rozpatrywane pod względem ich wpływu na poczucie realizmu prezentowanych obrazów.

Praca kończy się podsumowaniem i zaproponowaniem dalszych kierunków rozwoju w zakresie omawianej tematyki.

11 Zawartość pracy

Cała praca stanowi wieloaspektowe ujęcie problematyki głębi i realizmu obrazów, które powstają w wyniku komputerowej syntezy i są dostarczane odbiorcy za pomocą technik wyświetlania mających stymulować postrzeganie przestrzenności trójwymiarowych scen. W ramach rozprawy zostały opracowane metody, które wykorzystują pomiar kierunkowości widzenia człowieka, naśladując mechanizm akomodacji ludzkiego oka, posługując się obrazowaniem stereoskopowym i kontrastem w wyświetlaczach o poszerzonym zakresie dynamiki luminancji (tzw. *wyświetlacze HDR*).

11.1 Modelowanie kierunkowości widzenia

Uwaga wizualna jest zbiorem indywidualnych procesów kognitywnych, które odbywają się w sposób ukryty dla otoczenia. Możliwość ich poznania mogłaby dostarczyć ważnego wglądu w naturę i procesy myślowe człowieka. Jest to jedna z najistotniejszych myśli, które prowokowały badaczy do tego, by starać się ten wgląd uzyskać. Naturalnym więc jest, że przedmiotem badania stał się jedyny zewnętrzny objaw uwagi jakim jest kierunek spojrzenia, a więc orientacja gałek ocznych i ich ruch. Techniki obserwacji, śledzenia i interpretacji ruchów oczu nazywa się okulografią.

Źródłem danych dla okulografii jest pomiar ruchów oczu badanej osoby. Na podstawie zaobserwowanego stanu gałek ocznych wyznacza się kierunek spojrzenia, w celu późniejszego wnioskowania o aktualnym przedmiocie, na którym skupia się badany.

W przypadku uproszczonego założenia, że fizyczny kierunek patrzenia wskazuje na skupienie uwagi na obiekcie znajdującym się na osi wyznaczonej przez ten kierunek, powstaje problem wyboru obiektu spośród zbioru obiektów spełniających

to kryterium. Problem ten jest szczególnie wyraźny w środowiskach wirtualnej rzeczywistości, które zazwyczaj opierają się na syntezie obrazu trójwymiarowej sceny złożonej z licznych, niezależnych obiektów przestrzennych. W takim wypadku za obiekt uwagi najczęściej uznaje się ten nieprzezroczysty obiekt, który znajduje się najbliżej wirtualnej kamery, posiłkując się informacją odczytaną z bufora głębokości [MBT11]. Innym, słuszniejszym podejściem które można zastosować w sytuacji, gdy prezentujemy obserwatorowi obraz stereoskopowy, jest wybór obiektu znajdującego się w odległości związanej z płaszczyzną na której faktycznie skupiany jest wzrok, mierząc ruchy konwergencyjne oczu [DPHW11].

11.1.1 Problem jakości danych

Istotną kwestią, którą należy brać pod uwagę korzystając ze współczesnych systemów eye trackingowych, jest problem jakości danych jakie są przez nie dostarczane. Należy liczyć się z tym, że dane uzyskiwane z pomiarów obarczone są błędem w stosunku do rzeczywistego kierunku patrzenia użytkownika. A co za tym idzie, wnioskowanie o obiekcie, na którym skupiona jest uwaga obserwatora, może być jeszcze mniej dokładne.

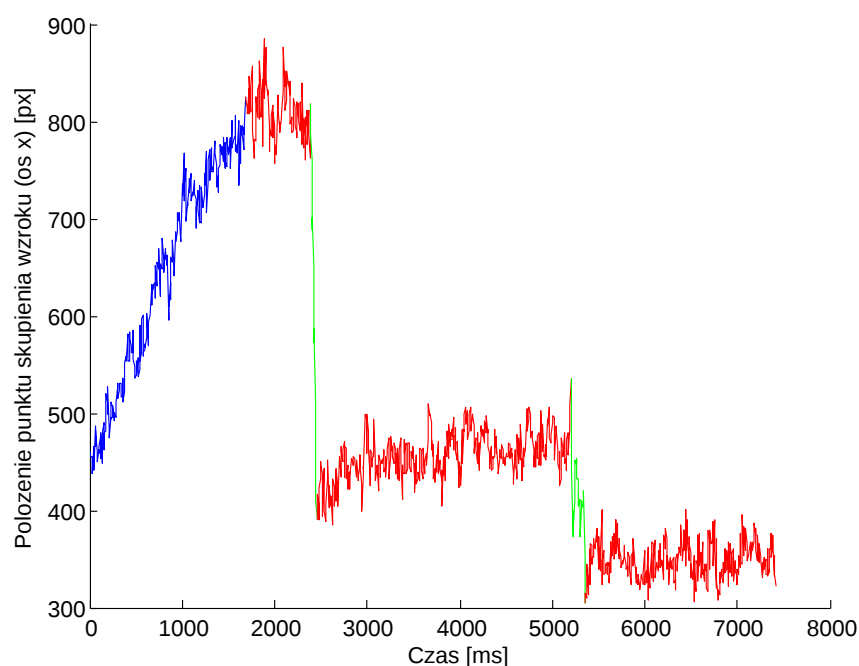
Współczesne urządzenia służące okulografii obarczone są istotnymi ograniczeniami technicznymi. Korzystając z metod opartych o obraz wideo gałek ocznych, nie jest możliwe zmierzenie kierunku patrzenia z idealną dokładnością. Decydujące są tutaj czynniki takie jak *rozdzielczość obrazu rejestrowanego przez kamerę, częstotliwość pracy kamery* czy w końcu jej *własności optyczne*. Także *model matematyczny* pozwalający na wyznaczenie kierunku na podstawie obrazów z kamery często oparty jest o znaczące uproszczenia, jak choćby założenie o jednorodności krzywizny powierzchni rogówki.

Wobec braku jednoznacznego sposobu oceny jakości rozwiązań okulograficznych, przyjęło się wyróżniać trzy istotne aspekty charakteryzujące dany system eyetrackingowy: precyzję (ang. *precision*), dokładność (ang. *accuracy*) i powtarzalność pomiarów (ang. *robustness*) [tob11, HNA⁺11].

Zaszumienie sygnału, będące cechą charakterystyczną danych obarczonych błędem precyzji, jak i systematyczne jego przesunięcie względem rzeczywistego kierunku spojrzenia związane z małą dokładnością, są zjawiskiem powszechnym w okulografii. Przykładowy sygnał otrzymany z urządzenia do śledzenia wzroku przedstawiono na Rysunku 1.

Aby możliwe było efektywne wykorzystanie takich danych w celu wnioskowania o uwadze badanej osoby, konieczne jest ich przekształcenie w celu zminimalizowania wpływu błędów na ostateczny rezultat śledzenia uwagi. Proste metody wygładzania sygnału takie jak średnia krocząca czy też średnia ważona, choć znacząco zmniejszają zaszumienie, to z uwagi na skokowy charakter zmian kierunku spojrzenia w momencie wystąpienia sakad prowadzą do dalszego zniekształcenia sygnału i odsunięcia wyliczonego punktu skupienia wzroku od rzeczywistego jego miejsca. Należy liczyć się także z opóźnieniem, jakie tego rodzaju metody wprowadzają przy zastosowaniach o charakterze *online*.

Dość naturalną koncepcją stało się klasyfikowanie fragmentów przebiegu czasowego związane z fizjologicznymi ruchami gałek ocznych. Najczęściej w ogólnych zastosowaniach okulograficznych bierze się pod uwagę przede wszystkim fiksacje oraz



Rysunek 1: Przykładowy przebieg czasowy współrzędnej x położenia punktu skupienia wzroku na ekranie. Kolorem czerwonym oznaczono przedziały zaklasyfikowane jako fiksacje, kolorem zielonym sakady, a kolorem niebieskim płynny pościg. Widać problem małej precyzji w postaci zaszumienia sygnału. Dane źródłowe pochodzą z sesji, podczas której użytkownik był proszony o śledzenie wyświetlonego znacznika który wymuszał poszczególne ruchy, stąd długi czas trwania fiksacji, nienaturalny przy swobodnej obserwacji.

sakady, pierwsze z nich przyjmując za czas świadomego percypowania obrazu, a drugie odfiltrowując jako niepowiązane z odbieraniem obrazu przez umysł momenty przejścia pomiędzy fiksacjami. To powszechne uproszczenie pomija zupełnie okresy płynnego pościgu, które zostają błędnie sklasyfikowane jako serie fiksacji przedzielonych sakadami, gdy zasięg przemieszczenia próbek wyczerpie już znamiona bycia tymi drugimi. Niemniej dla statycznych bodźców i rozpatrywania uwagi w obrębie nieruchomej płaszczyzny np. ekranu, taka koncepcja zdaje się być skuteczna. Sztandarowymi przykładami implementacji tej idei są algorytmy *I-DT* oraz *I-VT* [SG00]. Obie te metody mają na celu wyodrębnienie punktów fiksacji z dwuwymiarowego sygnału pochodzącego z okulografu.

Szczególnym przypadkiem dla filtracji danych okulograficznych jest sytuacja, w której możliwe jest wyznaczenie zbioru elementów sceny, które mogą stać się odbiorcami wizualnej uwagi obserwatora. Wówczas można rozpatrzyć uproszczenie problemu filtracji danych do dyskretnej postaci, w której oczekiwanym rezultatem jest wskazanie oglądanego elementu sceny, a nie konkretnego miejsca. Takie metody identyfikacji noszą zbiorczą nazwę *I-AOI* (ang. *Identification - Area Of Interest*) [SG00].

Natomiast mając do dyspozycji dane o punkcie skupienia wzroku określone w domenie ciągłej, wyznaczenie właściwego, odpowiadającego im elementu ze zbioru potencjalnych odbiorców uwagi, można sprowadzić do określenia kartezjańskiej odległości pomiędzy każdym z nich a punktem skupienia uwagi. Wówczas najbliższy z elementów jest wskazywany jako ten, z którym powiązana jest dana próbka, która mogła wcześniej zostać poddana filtracji z użyciem np. metody *I-DT* lub *I-VT*. Jest to metoda naiwna, okazująca swoje słabości szczególnie w przypadku złożonych wizualizacji dynamicznych środowisk trójwymiarowych, gdzie poruszające się obiekty znajdujące się na różnych planach mogą zostać z łatwością błędnie wskazane.

Identyfikacja obiektu będącego przedmiotem uwagi podczas prezentowania bodźców dynamicznych, wymaga wzięcia pod uwagę również płynnych pościgów (ang. *smooth pursuits*), wykonywanych przez oczy obserwatora w celu skupienia się na poruszających się elementach. Wykorzystanie rozpoznawania tego rodzaju ruchów jest tematem stosunkowo nowym, będącym głównie przedmiotem zainteresowania badaczy pracujących nad systemami rzeczywistości rozszerzonej (ang. *augmented reality*) bazujących na obrazie pochodzącym z otoczenia obserwatora [VBG13, KvI⁺16].

Podczas prac związanych z niniejszą rozprawą nie była znana żadna technika identyfikacji, która dedykowana byłaby systemom w których możliwe jest pozyskanie informacji o topologii oglądanej sceny. Dlatego istotnym elementem pracy stało się opracowanie takiego podejścia, które korzystałoby z dostępnej na bieżąco wiedzy o dynamicznej scenie w celu osiągnięcia jak najlepszego rezultatu minimalizującego błędy wskazywania odbiorcy uwagi. Tak powstała opisana niżej metoda *GDOT*, zorientowana na użycie przede wszystkim w środowiskach wirtualnej rzeczywistości.

11.1.2 Metoda GDOT

Zaproponowana w pracy technika *GDOT* (ang. *Gaze-Driven Object Tracking*) opiera się na zastosowaniu modelu probabilistycznego w celu wyznaczenia tego elementu sceny, który z największym prawdopodobieństwem jest w danej chwili oglądany przez obserwatora. Głównym założeniem tej techniki było połączenie danych do-

starczanych na bieżąco przez eye tracker z wiedzą o oglądanej scenie, jak i z wiedzą o charakterystycznych cechach ludzkiego widzenia. W ten sposób powstała metoda pozwalająca na znaczącą poprawę skuteczności identyfikacji obiektu uwagi, ukierunkowana szczególnie na działanie w warunkach dużej zmienności sceny. To podejście, opracowane w ramach pracy nad rozprawą, zostało opublikowane we współautorskim artykule [MBM13].

Zasada działania

Nie ulega wątpliwości, że aby obiekt został wskazany jako ten, na którym spoczywa wzrok oglądającego, pomierzone położenie punktu skupienia wzroku musi być podobne do położenia obiektu. Jest to punkt wyjściowy znanych metod filtracji danych, przedstawionych w poprzednim rozdziale. Także dla metody *GDOT* jest to jedną z dwu podstawowych przesłanek stanowiących o prawdopodobieństwie śledzenia danego obiektu przez obserwatora.

Kolejnym, ważnym założeniem stanowiącym podstawę idei działania algorytmu, jest powiązanie śledzenia wzrokiem ruchomych obiektów z inicjalizacją tego ruchu gałek ocznych, który nazwany został płynnym pościgiem. Czyniąc takie założenie można spodziewać się, że sygnał położenia punktu skupienia wzroku obserwatora w czasie, będzie skorelowany z sygnałem położenia śledzonego obiektu. To zaś przekłada się na oczekiwane powiązanie pochodnych sygnałów położenia, a więc prędkości przemieszczania się obu punktów.

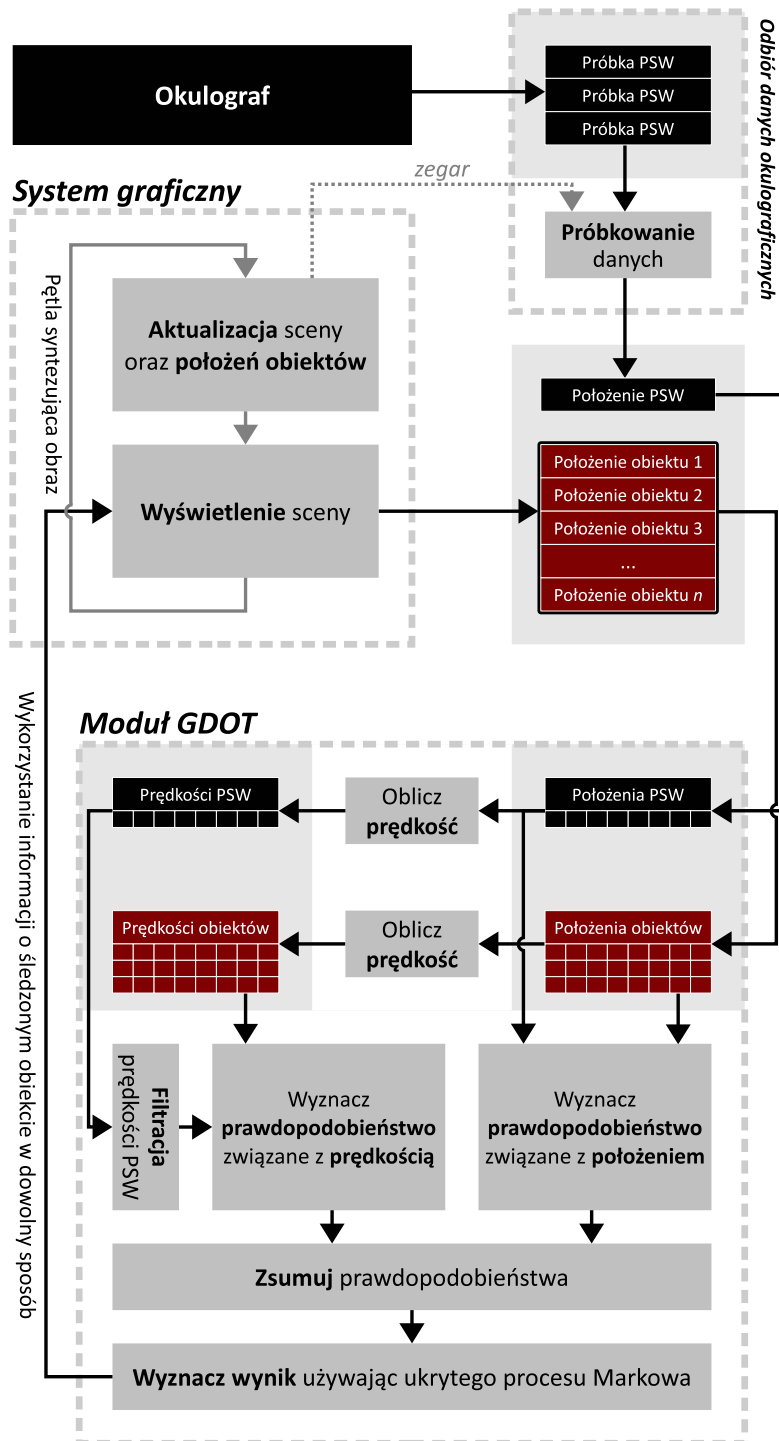
GDOT operuje na skończonym, predefiniowanym zbiorze obiektów. Wyjściem algorytmu jest dyskretna wartość będąca wybranym, najbardziej prawdopodobnym elementem z tego zbioru. Prawdopodobieństwo obliczane jest na podstawie porównania położenia i prędkości próbek pochodzących z okulografu, z próbkami zawierającymi dane o elementach sceny, pochodzącymi z systemu prezentującego obserwatorowi oglądany obraz. Dlatego ważnym jest, by system mający wykorzystać algorytm *GDOT* był w stanie dostarczać na wejście algorytmu bieżące informacje o położeniach każdego z obiektów mogących stać się potencjalnym odbiorcą uwagi użytkownika.

Ocena jakości

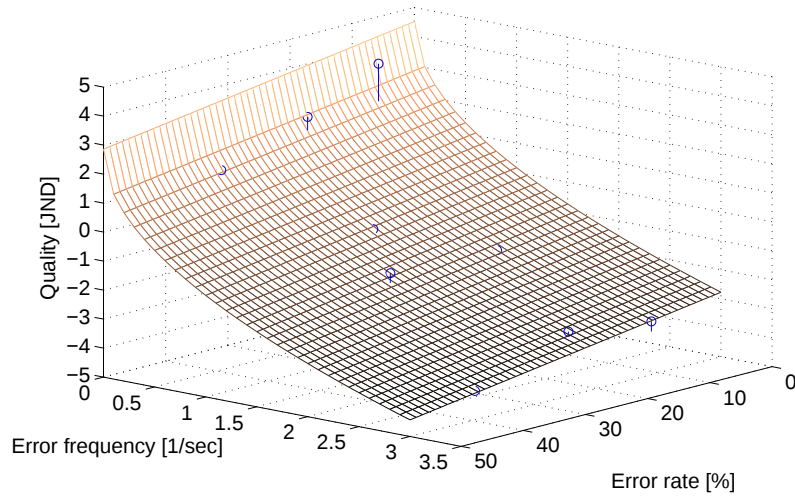
Do porównania skuteczności identyfikacji obiektu uwagi w oparciu o dane okulograficzne wykorzystano opracowaną w ramach rozprawy metrykę Q , na którą składają się dwa komponenty: *udział błędu identyfikacji* E_{rate} wyrażony jako stosunek czasu, przez który wskazywany jest błędny obiekt do całości czasu trwania sesji oraz *częstotliwość błędnej identyfikacji* E_{freq} wyrażona liczbą zmian wskazywanego obiektu na błędny w ciągu jednej sekundy. Dobra technika poprawy jakości danych okulograficznych powinna dążyć do minimalizacji obu tych czynników, jednak ich wagi opierają się o subiektywne odczucia obserwatorów, dlatego zostały one dobrane na podstawie przeprowadzonego eksperymentu percepcyjnego. Po dokonaniu aproksymacji uzyskanych rezultatów uzyskano równanie powierzchni prezentowanej na Rysunku 3:

$$Q = -0.03312 \cdot E_{rate} - 4.358 \cdot E_{freq}^{0.4682} + 4.516.$$

Rysunek 2: Schematyczne przedstawienie sposobu działania algorytmu identyfikacji obiektu uwagi *GDOT*. Skrótem *PSW* oznaczono Punkt Skupienia Wzroku.



Rysunek 3: Powierzchnia powstała w wyniku dopasowania do rezultatów uzyskanych z eksperymentu, który posłużył do zbudowania metryki jakości.



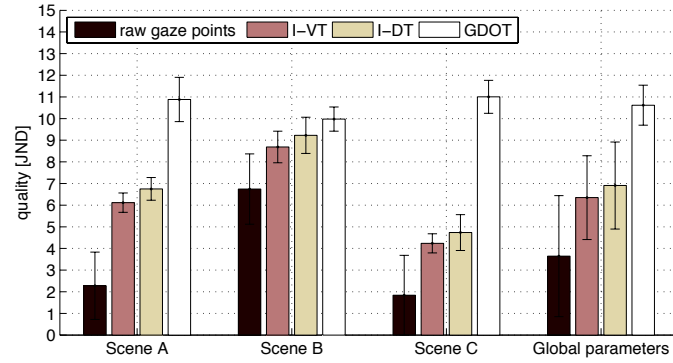
Uzyskana metryka faworyzuje sytuacje, w których częstotliwość zmian na błędne wskazanie jest najmniejsza.

Eksperyment oceny jakości działania metody *GDOT* polegał na prezentacji obserwatorom trzech animacji zsyntezowanych z użyciem technik grafiki komputerowej czasu rzeczywistego. Wszystkie trzy animacje związane były z jedną, trójwymiarową sceną przedstawiającą kilka obiektów poruszających się wewnątrz sześciennego pomieszczenia. Sekwencje różniły się charakterem oraz intensywnością ruchów obiektów i kamery. Przez cały czas ekspozycji sekwencji wideo na ekranie zaznaczony był punkt, o dokładne śledzenie którego proszono uczestników. Punkt ten podążał za poszczególnymi obiektami lub ich fragmentami oraz przemieszczał się pomiędzy nimi, mając za zadanie imitować ukierunkowanie uwagi osoby oglądającej scenę. W eksperymencie założono, że obserwator patrzy na ten punkt, a więc to właśnie położenie tego punktu powinno być identyfikowane przez algorytm detekcji fiksacji.

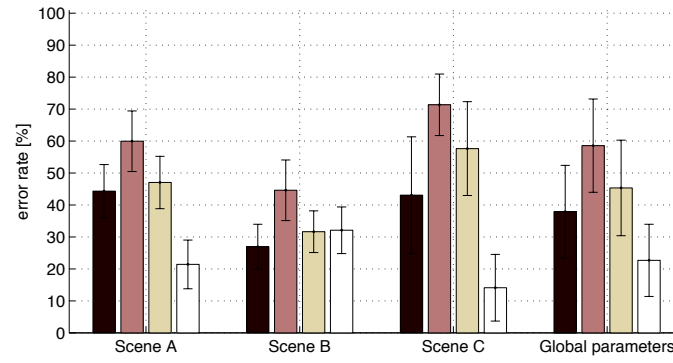
Dla testowego zbioru danych uzyskano wyniki, które wskazują na znacznie lepszą jakość identyfikacji obiektów z użyciem metody *GDOT*, niż uzyskuje się korzystając z innych metod (zob. Rys. 4a).

Cechą szczególną metody *GDOT* jest znaczne ograniczenie częstotliwości przechodzenia do wskazywania błędnego obiektu. Dążenie do niezmienności wyjścia algorytmu wiąże się czasem z podtrzymywaniem błędnego wyniku identyfikacji na rzecz zredukowania liczby przejść pomiędzy stanami, w wyniku czego procentowy udział błędu może ulec nawet nieznacznemu zwiększeniu. Jednak zgodnie z wyznaczoną i opisaną wcześniej percepcyjną metryką jakości, taka zależność może być korzystna z punktu widzenia użytkownika aplikacji wykorzystującej dane pochodzące z eye trackera. To zresztą obrazuje ostateczny rezultat na Rys. 4a.

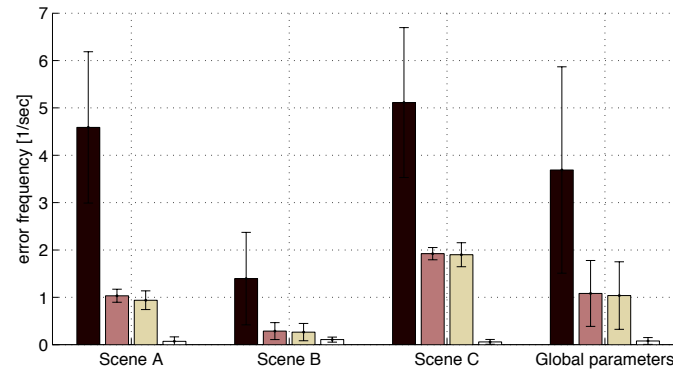
Na Rys. 5 przedstawiony jest przykładowy przebieg czasowy identyfikacji obiektów uwagi dla różnych metod filtracji danych. W idealnym przypadku linia niebieska, będąca zobrazowaniem wyjścia identyfikacji obiektu uwagi, powinna pokrywać się z linią czerwoną oznaczającą ten obiekt, którego śledzenie było zadaniem obserwatora. Metoda *GDOT* najlepiej odwzorowuje sygnał referencyjny. Niewielki udział



(a)

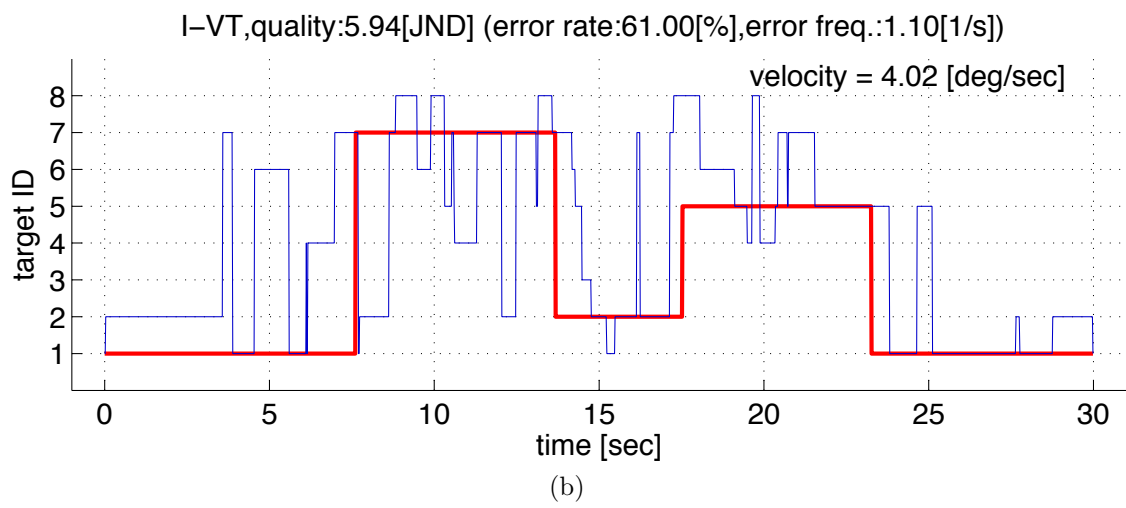
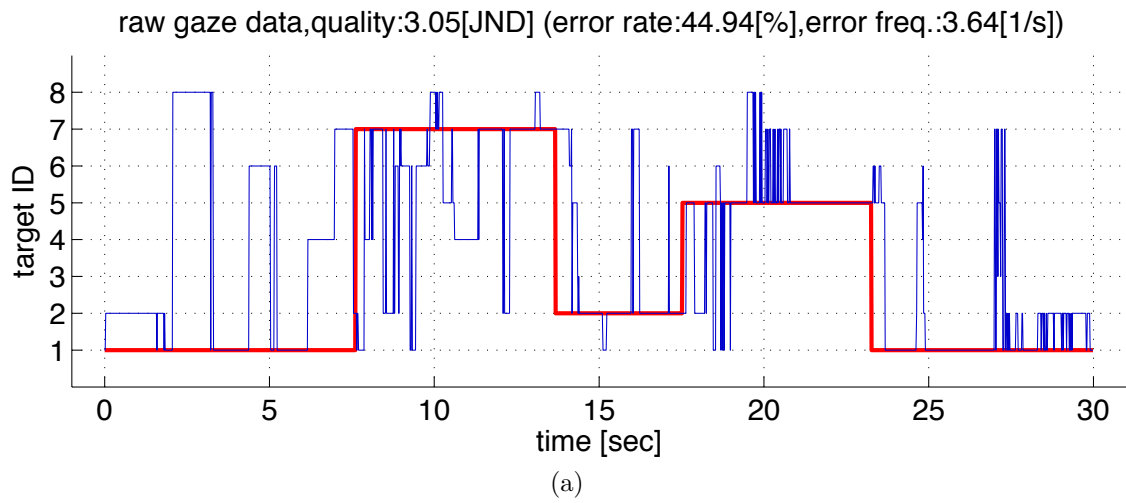


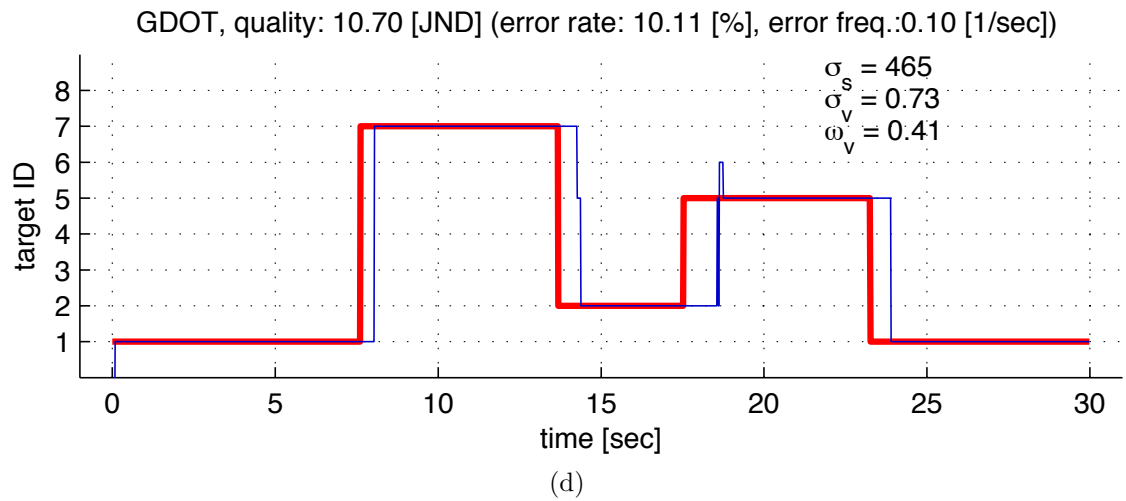
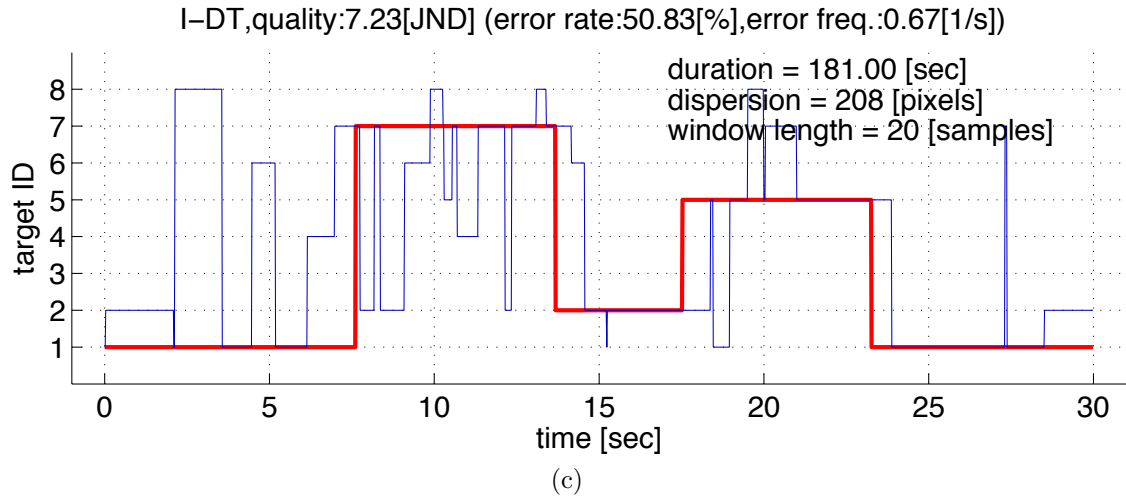
(b)



(c)

Rysunek 4: Jakość identyfikacji według opisanej w rozprawie metryki (a) oraz dwie składowe: udział błędu (b) i jego częstotliwość (c), z podziałem na poszczególne animacje oraz dla przypadku gdy użyto uniwersalnie zoptymalizowanych wartości parametrów. Słupki koloru czarnego pokazują surowe dane okulograficzne, kolorem różowym oznaczono metodę *I-VT*, kolorem żółtym metodę *I-DT*, natomiast słupki koloru białego pokazują wyniki dla proponowanej metody *GDOT*.





Rysunek 5: Przykładowe przebiegi czasowe identyfikacji obiektu uwagi z użyciem różnych metod filtracji dla tego samego zestawu danych okulograficznych pochodzącego od jednego obserwatora. Kolejno: surowe dane (a), metoda *I-VT* (b), metoda *I-DT* (c) i metoda *GDOT* (d). Pozioma oś to czas, natomiast na pionowej osi naniesiono identyfikatory obiektów z puli, która zawierała preselekcjonowany zbiór potencjalnych odbiorców uwagi. Kolorem czerwonym zobrazowano sygnał referencyjny, a więc obiekt o śledzenie którego proszony był obserwator. Ilustracja zaczerpnięta z autorskiej publikacji [BM14a].

błędu wywołany jest przez opóźnienie wynikające z techniki radzenia sobie z podtrzymaniem ciągłego wskazywania już raz wybranego obiektu. Dzięki zachowawczości oraz uwzględnianiu kierunku przemieszczania się obiektów i punktu skupienia wzroku oraz rozkładu błędu, metoda *GDOT* może nie tylko doprowadzić do znaczącej redukcji częstotliwości błędu, ale też pozwala skutecznie wykrywać najbardziej prawdopodobny obiekt będący przedmiotem uwagi obserwatora.

11.1.3 Zastosowania metody *GDOT*

W rozprawie opisano trzy sposoby wykorzystania metody *GDOT*: identyfikacja przedmiotu zainteresowania na potrzeby analizy uwagi osób oglądających wideo [BM14a], przyspieszenie syntezy obrazu poprzez redukcję jego szczegółowości w częściach nieobjętych szczegółowym widzeniem użytkownika [TB10] i projekt rozszerzeń klasycznych elementów interfejsu użytkownika w grze komputerowej, opartych o wiedzę o oglądanym w danej chwili obiekcie `citeownbazyluk14sn`.

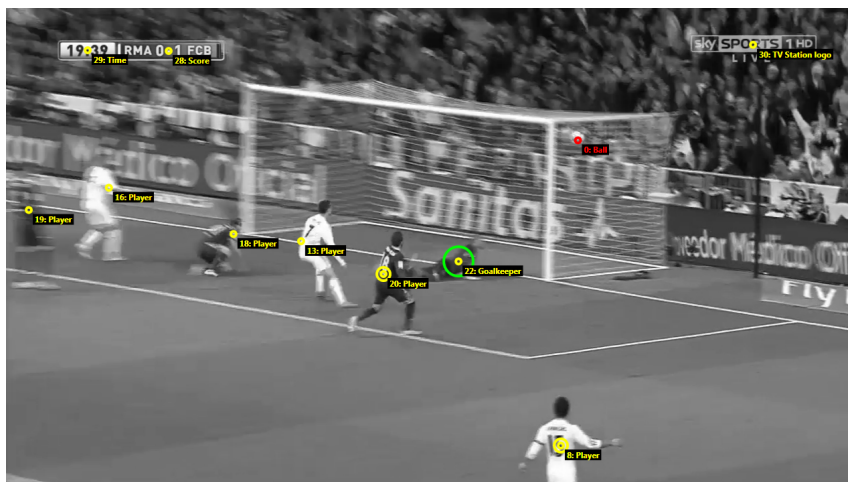
Analiza uwagi osób oglądających wideo

Sztandarowym przykładem sytuacji, w której zadaniem jest wybór jednego z widocznych na obrazie obiektów na podstawie miejsca w które spogląda widz, jest analiza przebiegu uwagi wizualnej w obrazie wideo. Sekwencje wideo obrazujące jakiś wycinek rzeczywistości przedstawiają najczęściej semantycznie rozróżnialne przedmioty i postacie, które mogą być traktowane jako całości w przypadku badania częstości i kolejności ich obserwacji przez widzów czy też szeroko pojętego zainteresowania nimi. Innymi słowy, pole to stanowi obszar o cechach zbliżonych do tych, które były podstawą stworzenia metody *GDOT*, a tym samym jest to miejsce gdzie ta metoda może znaleźć swoje zastosowanie.

W proponowanym rozwiązaniu ekspert, osoba przygotowująca materiał do badania, korzystając ze swojej wiedzy o semantyce prezentowanego obrazu, jest odpowiedzialny za wskazanie w specjalnie przygotowanej aplikacji miejsc na obrazie w których znajdują się obiekty interesujące z punktu widzenia badania. W przeciwieństwie do klasycznych rozwiązań, nie musi on jednak ponownie, w ręczny sposób kontynuować wskazywania nowych położeń wybranych przez siebie elementów we wszystkich kolejnych klatkach sekwencji wideo. W zamian wykorzystywana jest technika analizy rzadkiego przepływu optycznego (ang. *sparse optical flow*) z użyciem metody Lucasa-Kanade [LK81], która pozwala w sposób zautomatyzowany śledzić poruszanie się fragmentu sceny w czasie.

Znając już skuteczność algorytmu *GDOT* dla sekwencji wideo syntezowanych komputerowo, gdzie położenie potencjalnych obiektów uwagi jest precyzyjnie wyznaczane na podstawie danych z silnika renderującego, konieczne było przeprowadzenie eksperymentu potwierdzającego tą skuteczność w opisanym wyżej przypadku, gdy informacje o położeniu pochodzą z niedokładnego śledzenia metodą opartą o wyznaczenie przepływu optycznego. Za przykład sekwencji wideo do testów obrano fragmenty telewizyjnej transmisji meczu piłki nożnej. Charakter tego obrazu jest bardzo interesującym przypadkiem dla postawionego zadania, ponieważ możliwe jest logicznie uzasadnione wyznaczenie licznego zbioru obiektów sceny, które są w ciągłym ruchu z różnymi prędkościami i mają częstą tendencję do wzajemnego

Rysunek 6: Przykładowa klatka z testowych sekwencji wideo pochodzących z telewizyjnej transmisji meczu piłki nożnej (źródło: *SkySports*) z zaznaczonymi elementami zbioru potencjalnych odbiorców uwagi. Ich wizualizacja pochodzi z autorskiego oprogramowania służącego do wyznaczania tych obiektów.



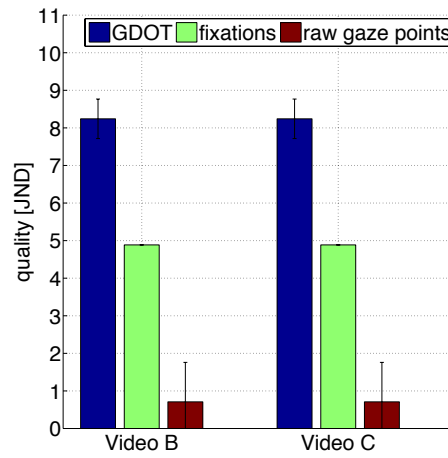
przesłaniania się. Stąd naturalnym jest, że wyznaczonymi elementami zbioru potencjalnych odbiorców uwagi byli przede wszystkim piłkarze, ale także piłka, sędziowie i statyczne elementy takie jak prezentowany na ekranie wynik lub czas trwania spotkania (por. Rys. 6). W badaniu użyto trzech krótkich sekwencji (trwających odpowiednio 18, 29 i 13 sekund), nazwanych kolejno *A*, *B* i *C*.

Uzyskane przebiegi punktów skupienia wzroku poddano działaniu detekcji fiksacji w oparciu o powszechnie stosowaną metodę *I-DT*. Porównano wyniki identyfikacji obiektów na podstawie tak pozyskanych danych dla trzech przypadków: gdy źródłem były surowe dane, gdy identyfikację oparto na fiksacjach oraz gdy za identyfikację w całości odpowiadał algorytm *GDOT* (zob. Rys 8). Dla sekwencji *B* i *C*, gdzie możliwe było wskazanie sygnału referencyjnego (dla *B* obiekt oznaczony w danej chwili markerem, dla *C* piłka), porównano uzyskane wyniki stosując metrykę opisaną wcześniej w tym rozdziale. Rezultat zobrazowano na Rys. 7.

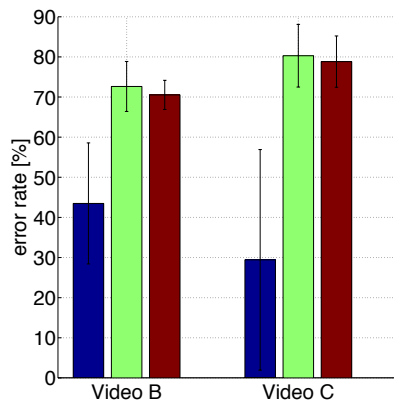
Podobnie jak w przypadku, gdy źródłem obrazu i informacji o położeniu obiektów sceny był system grafiki czasu rzeczywistego, także i tu da się zauważyć wyraźną przewagę jakości identyfikacji obiektu uwagi nad pozostałymi przypadkami, gdy wykorzystywany jest algorytm *GDOT*.

Przyspieszenie syntezy obrazu

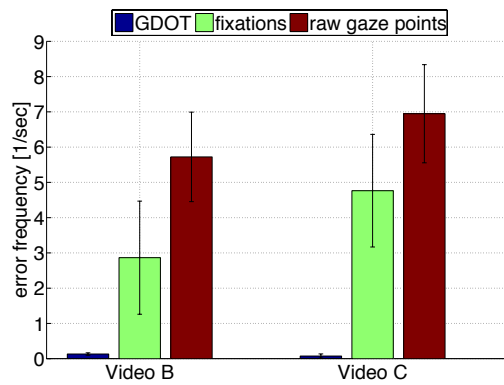
Obliczenia związane ze złożonymi efektami graficznymi stosowanymi do już zsyntezowanych klatek (ang. *post processing*) wymagają dużej ilości zasobów. Tego rodzaju operacje mają zazwyczaj charakter *SIMD* (ang. *Single Instruction Multiple Data*). Oznacza to, że najczęściej ich złożoność związana jest z liczbą powtórzeń zestawu instrukcji realizującego dany efekt, a więc liczbą pikseli koniecznych do przetworzenia, co z kolei związane jest z rozdzielczością bufora klatki (dla operacji per-fragment realizowanych na geometrii sceny, liczba powtórzeń może być większa od liczby pikseli bufora klatki). Biorąc pod uwagę szerokie ujęcie systemu graficznego, obejmujące



(a)

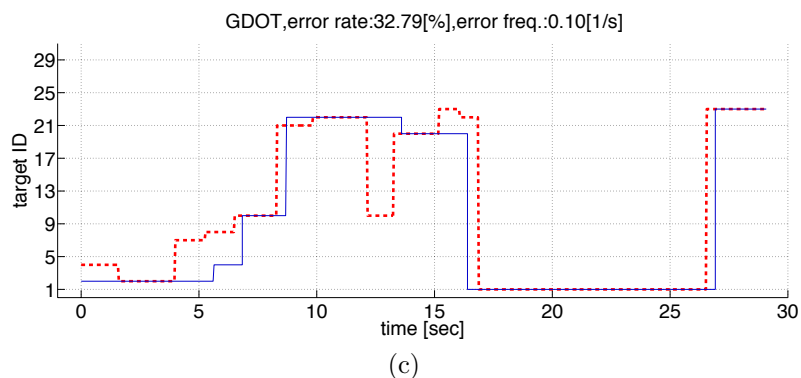
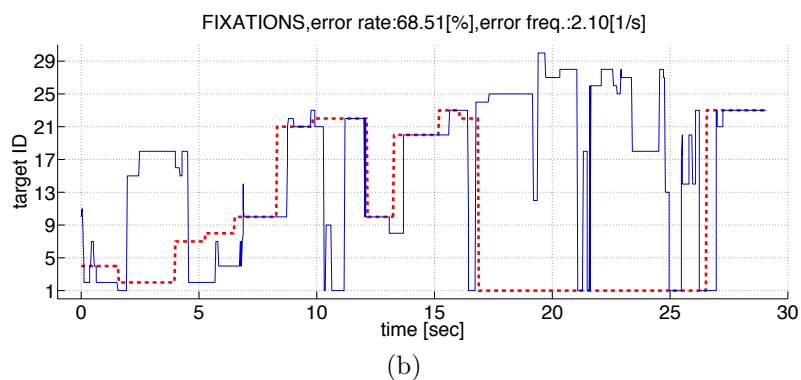
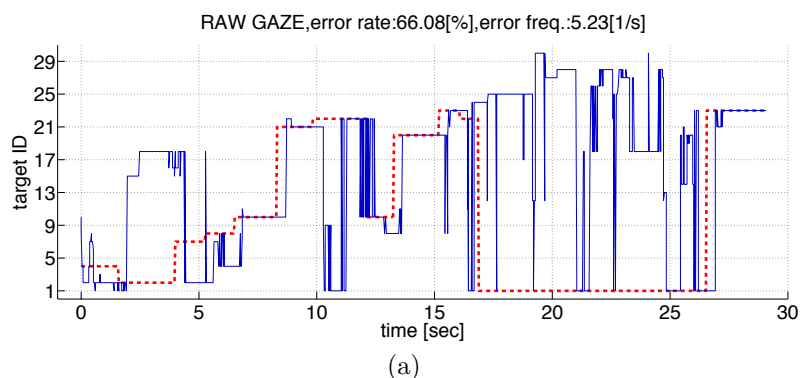


(b)



(c)

Rysunek 7: Jakość identyfikacji według opisaney w rozprawie metryki (a) oraz dwie składowe: udział błędu (b) i jego częstotliwość (c), z podziałem na poszczególne sekwencje *B* i *C*. Ilustracja zaczerpnięta z autorskiej publikacji [BM14a].



Rysunek 8: Przykładowe przebiegi czasowe identyfikacji obiektu uwagi w sekwencji B z użyciem różnych metod filtracji, uzyskane dla tego samego zestawu danych okulo-graficznych pochodzącego od jednego obserwatora. Kolejno: surowe dane (a), detekcja fiksacji z użyciem $I-DT$ (b), metoda $GDOT$ (c). Pozioma oś to czas, natomiast na pionowej osi naniesiono identyfikatory obiektów z puli, która zawierała preselekcjonowany zbiór potencjalnych odbiorców uwagi. Kolorem czerwonym zobrazowano sygnał referencyjny, a więc obiekt o śledzenie którego proszony był obserwator.

także użytkownika, nie można pominąć aspektów związanych z ograniczeniami jego aparatu widzenia. Mając na względzie niejednorodną dystrybucję fotoreceptorów w siatkówce oka, a co za tym idzie ograniczone, centralne pole widzenia szczegółowego, łatwo dojść do wniosku że w przypadku ekranu oglądanego tylko przez jednego widza, spora część obrazu jest w danej chwili objęta jedynie niedokładnym widzeniem peryferyjnym. Tym samym można rozważyć zaoszczędzenie zasobów wykorzystywanych dla przetworzenia tych pikseli, które nie są w danej chwili wyraźnie widziane przez obserwatora.

W ten sposób wprowadzone zostaje pojęcie obszaru zainteresowania nazywanego w skrócie *ROI* (ang. *Region Of Interest*), a więc tego wycinka obrazu prezentowanego widzowi, który jest w danej chwili oglądany za pomocą szczegółowego widzenia centralnego. W rozprawie zaproponowano próbę przyspieszenia w ten sposób symulacji zjawiska głębi ostrości, która odbywa się metodą opisaną dalej. To autorskie podejście zostało opublikowane w artykule [TB10].

Przyjęto, że obszar poza *ROI* będzie renderowany do bufora klatki o rozdzielczości dwukrotnie mniejszej wzdłuż każdego z boków niż docelowa rozdzielczość finalnej klatki. Zawartość bufora o pomniejszonej rozdzielczości będzie następnie wyświetlana na ekranie z użyciem filtracji dwuliniowej. Tym samym obszar poza *ROI* sprawia wrażenie rozmytego, niskoczęstotliwościowego w dziedzinie kontrastu, aby zredukować ryzyko zauważenia przez obserwatora zmian wysokiej częstotliwości w polu jego widzenia peryferyjnego [RFN92]. Faktyczna symulacja zjawiska głębi ostrości odbywa się więc tylko w obszarze *ROI*.

Badanie przyspieszenia, objawiającego się zwiększoną liczbą klatek na sekundę, przeprowadzono na komputerze wyposażonym w kartę graficzną *NVIDIA GeForce 8600M GT*. Scena złożona była z 21891 wierzchołków, a pełna klatka miała rozdzielczość 800×600 pikseli (zob. Rys. 9). Rezultaty przedstawiono na Rysunku 10.

Okazuje się, że zakładając kątowny rozmiar *ROI* z punktu widzenia obserwatora na poziomie 15° , przy ekranie o przekątnej 15 cali oglądanym z odległości około 50 cm, udaje się uzyskać wzrost liczby klatek na sekundę na poziomie 20% w stosunku do sytuacji, w której efekt jest stosowany do całej klatki.

Interfejsy użytkownika oparte na danych okulograficznych

Osobną gałęzią zastosowań okuloigrafii w graficznych systemach komputerowych jest wykorzystanie pozyskanej w procesie ciągłej obserwacji spojrzenia użytkownika wiedzy do stworzenia nowych lub ulepszenia już istniejących metod interakcji. Jest to kolejny przykład zastosowania o charakterze *online*, gdzie zbierane dane wykorzystywane są w czasie zbliżonym do rzeczywistego.

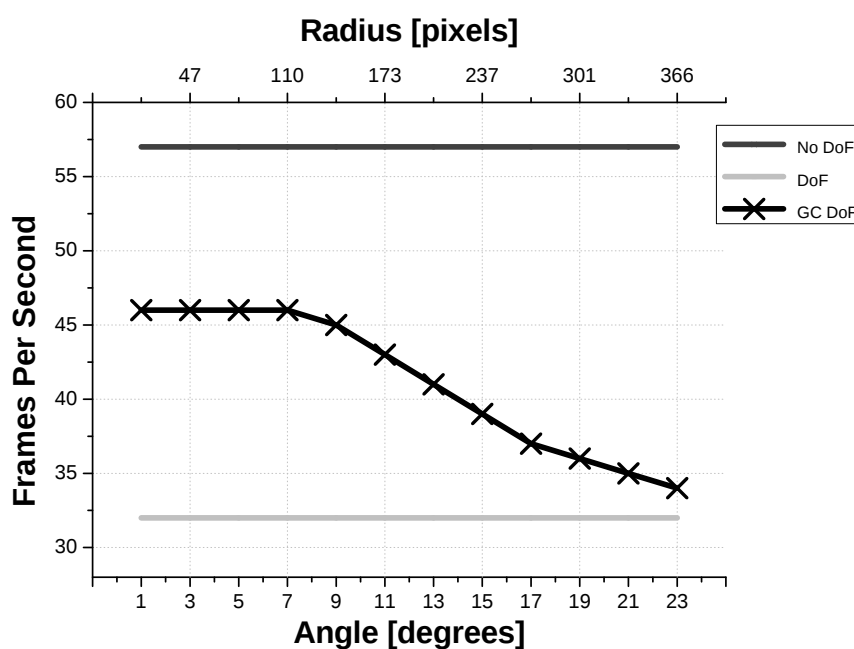
Przyjmując autorską, zręcznościową grę komputerową pt. *Invasion* (zob. Rys. 11) za środowisko badawcze, wprowadzono do niej trzy elementy interakcji wykorzystujące dane pozyskane z pomocą okuloigrafii:

- Kontekstowe informacje o oglądanych przez gracza obiektach, wyświetlane w sposób reagujący na aktualne miejsce skupienia jego uwagi
- Ruch wirtualnej kamery służącej do obrazowania wycinka trójwymiarowego świata na ekranie monitora realizuje automatyczne śledzenie wskazanego spoj-

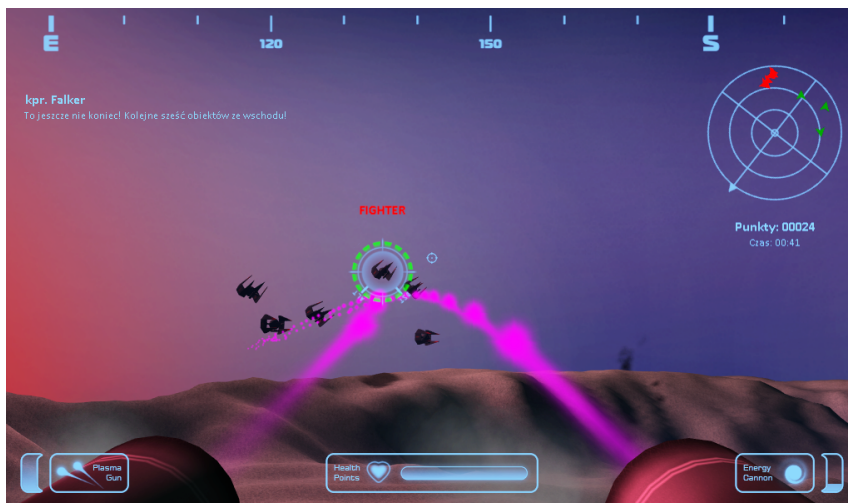
Rysunek 9: Klatka z wizualizacją sceny, która posłużyła ocenie przyrostu liczby klatek na sekundę związanej z rozmiarem *ROI*. Widoczna jest symulacja głębi ostrości ograniczona jedynie do *ROI* (obszar wokół zaznaczonych punktów: surowego skupienia wzroku - czerwony, fiksacji - zielony). Płynne zmieszanie krawędzi obszaru *ROI* z resztą klatki o obniżonej rozdzielczości, gdzie głębia ostrości nie jest symulowana.



Rysunek 10: Porównanie liczby klatek na sekundę podczas syntezy testowej klatki bez symulacji zjawiska głębi ostrości (czarna linia), z symulacją (szara linia) oraz z symulacją opartą o *ROI* (czarna linia z krzyżykami). Na poziomej osi oznaczono kątową miarę przeliczaną na długość boku *ROI* na ekranie. Ilustracja zaczerpnięta z autorskiej publikacji [TB10]



Rysunek 11: Autorska gra komputerowa *Invasion*, która służy jako podstawa do przeprowadzanych badań.



rzeniem przeciwnika, ułatwiając tym samym celne oddawanie strzałów, będące głównym elementem rozgrywki

- Występuje interaktywna symulacja zjawiska głębi ostrości, mającego imitować wrażenia wizualne znane z rzeczywistych układów optycznych

Są to autorskie rozwiązania opublikowane w artykule [BM14b]. Przeprowadzono badanie mające na celu potwierdzenie hipotezy o pozytywnym wpływie proponowanych elementów interfejsu użytkownika na uczucie zanurzenia i atrakcyjności rozgrywki. Jako że są to miary czysto subiektywne, zdecydowano o pozyskaniu wyników w drodze ankietyzacji grupy osób, której zostaną przedstawione do porównania dwie wersje tej samej gry.

Udało się potwierdzić, że zbierane w czasie rzeczywistym dane o punkcie skupienia wzroku gracza mogą posłużyć do zwiększenia atrakcyjności rozgrywki oraz zwiększenia poczucia zanurzenia gracza w wirtualnym świecie. Proponowane sposoby wzbogacenia zakresu interakcji z grą komputerową, która była przedmiotem badań, spotkały się z przychylnym odbiorem ze strony ankietowanych osób.

Wbrew założeniom, nie udało się jednak wykazać istotnego wpływu selektywnej prezentacji kontekstowej informacji wizualnej na jej czytelność. W tym zakresie powinny zostać przeprowadzone dalsze badania z racji szczególnego charakteru i dynamiki tej aplikacji. Przedmiotem dalszych badań powinny być także inne formy wykorzystania danych okulograficznych w środowiskach wirtualnych, takie jak inne efekty wizualne (np. adaptacyjne mapowanie tonów dla obrazów o poszerzonym zakresie dynamiki luminancji, tzw. *HDR*) czy interakcji ze środowiskiem (np. reakcja elementów świata na spojrzenia).

11.2 Interaktywna symulacja głębi ostrości

Modelowanie widzenia głębi to takie świadome kontrolowanie zawartych w prezentowanym obrazie wskazówek o głębi, by wywołać u obserwatora zamierzone wrażenie

trójwymiarowości sceny prezentowanej na płaskim medium. Celem takich zabiegów może być wzbogacenie informacji umieszczonej na obrazie, zwiększenie jego realizmu czy też zredukowanie niedogodności związanych z wykorzystywaną techniką wyświetlania. Posłużyć się tu można wszelkimi wskazówkami głębi, zarówno mono-, jak i binokularnymi.

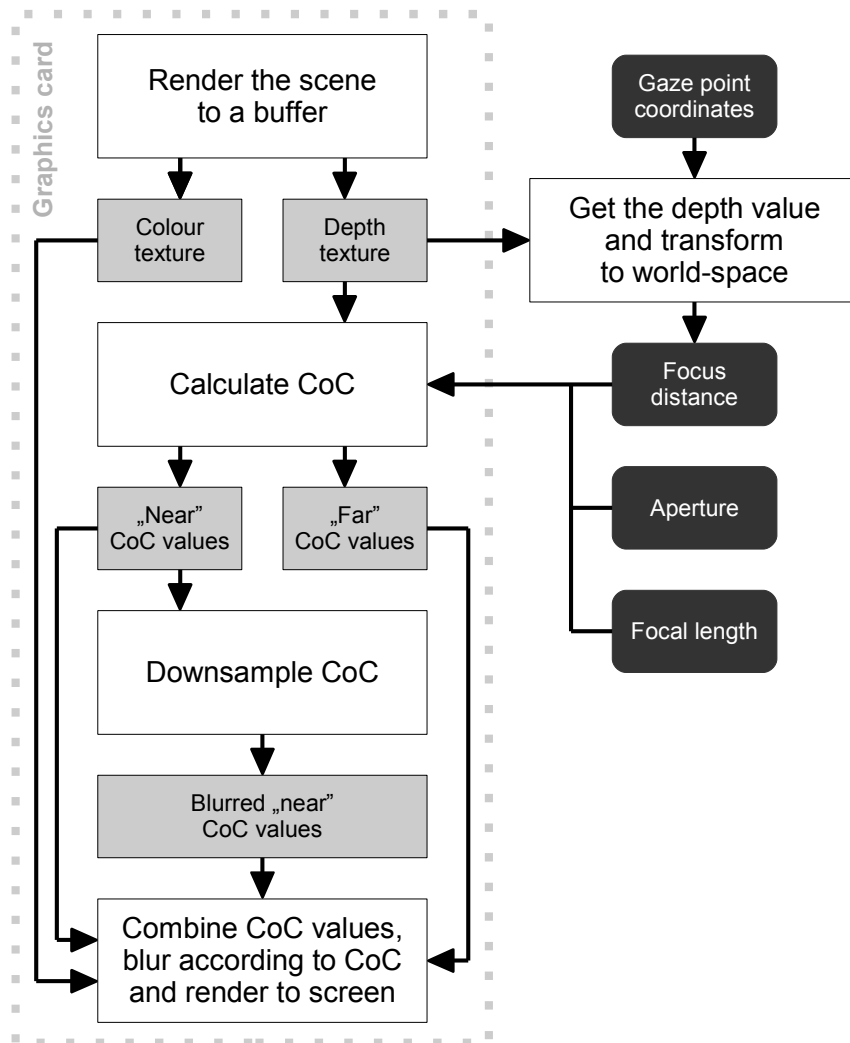
Głębia ostrości jest powszechnie spotykanym zjawiskiem wizualnym. Dotyczy ona każdego systemu optycznego wyposażonego w przynajmniej jedną soczewkę o niezerowych wymiarach, a więc zarówno urządzeń którymi posługujemy się w celu przechwytywania obrazu, jak i samego ludzkiego narządu wzroku. Będąc istotną, jednooczną wskazówką odnośnie głębi sceny, rozmycie obrazu wynikające z tego zjawiska jest często wykorzystywane w sposób przemyślany: zarówno dla uzyskania efektu artystycznego, wzmocnienia wrażenia głębi, jak i dla zamierzonego kierowania uwagą odbiorcy danej informacji wizualnej. Często jest celowo używana w fotografii i kinematografii, gdzie pozwala na zasygnalizowanie oglądającemu różnic w odległościach oraz wskazanie mu tych elementów sceny, które mają być główną treścią przekazu. Jednak w takich przypadkach widz godzi się na to, że rozmycie powstało już w momencie rejestrowania obrazu przez kamerę lub aparat fotograficzny. Jest więc wskazówką głębi zgodną z warunkami w tamtej chwili i odpowiadającą ustawieniu ostrości wybranemu przez operatora lub fotografa. Nie jest to więc wskazówka, która pozwalałaby swobodnie eksplorować scenę ukazaną na obrazie oczekując, że będzie się ona dostosowywać do intencji widza.

11.2.1 Implementacja i ewaluacja rezultatów

Aby pogodzić ze sobą zawarcie w obrazie informacji o głębi sceny w postaci rozmycia wynikającego z głębi ostrości, wraz z zaoferowaniem możliwości swobodnego wybierania przez obserwatora obiektu uwagi, opracowano koncepcję interaktywnej symulacji głębi ostrości (zob. Rys. 12). Idea ta polega na sztucznym dodaniu rozmycia do syntezywanego komputerowo obrazu naśladując to rozmycie, które wynikałoby z zachowania rzeczywistych odległości w obrazowanej scenie, gdy widz spoglądałby na ten sam element, na który spogląda widząc dwuwymiarową projekcję prezentowaną na ekranie. Dla pozyskania wiedzy o elemencie obrazu na którym w danej chwili skupia wzrok oglądający wykorzystywany jest okulograf [Baz10, MBT11].

Implementacja w systemie graficznym interaktywnej symulacji głębi ostrości nie tylko dostarcza obserwatorowi komfortu dowolności. Ma również za zadanie czynić oglądanie sceny znacznie bardziej naturalnym, zbliżonym do obserwowania rzeczywistego otoczenia, gdy skutkiem występowania zjawiska głębi ostrości jest rozmycie wynikające z akomodacji samego oka. Tym samym koncepcja ta ma przyczynić się do zwiększenia poczucia zanurzenia w prezentowanym na obrazie wirtualnym świecie, poprzez zwiększenie jego wiarygodności [Fer03, CF10]. Podobne rozwiązanie, stosowanie w systemach wirtualnej lub rozszerzonej rzeczywistości opartych na wyświetlaczach znajdujących się w pobliżu oczu, może służyć ułatwieniu akomodacji do różnych płaszczyzn [PKS⁺17].

Dla symulacji głębi ostrości użyto podejścia typu *post processing* i metody gromadzącej. Zastosowano renderowanie kolistego kształtu krążka rozmycia (*bokeh*), jako że w przypadku akomodacji człowieka za kształt ten odpowiada źrenica, a nie mechaniczna przysłona. Posłużono się przy tym próbkami z dysku Poissona o rów-



Rysunek 12: Schemat przepływu danych w systemie interaktywnej symulacji zjawiska głębi ostrości. Ilustracja zaczerpnięta z autorskiej publikacji [MBT11]



(a)



(b)

Rysunek 13: Porównanie tego samego ujęcia sceny pochodzącej z testowej implementacji pt. *Magician's House Demo* z wyłączoną (a) i włączoną (b) symulacją głębi ostrości wykorzystującą opisany algorytm. W drugim przypadku ostrość ustawiona jest na miejsce wskazane zielonym znacznikiem, które odpowiada współrzędnym punktu skupienia wzroku użytkownika.

nych wagach. Implementację oparto o programowalny potok renderowania. Kolejne etapy symulacji głębi ostrości realizowano jako programy cieniujące (ang. *shading programs*). Całość została zrealizowana w formie aplikacji *Microsoft .NET 3.5* napisanej w języku *C#*, wykorzystującej programowalny potok renderowania za pośrednictwem *OpenGL 3.2* (zob. Rys. 13).

System interaktywnej głębi ostrości utworzono na potrzeby weryfikacji hipotezy mówiącej, że interaktywna głębia ostrości zwiększa wrażenie realizmu prezentowanej wizualizacji trójwymiarowej sceny, jest więc preferowana przez obserwatorów. Badanie to opisano w publikacji, która powstała w ramach prac nad niniejszą rozprawą [MBT11].

Analiza wyników uzyskanych w eksperymencie pozwala na stwierdzenie, że uczestnicy badania zauważają różnicę pomiędzy interaktywną, a nieinteraktywną wersją symulacji zjawiska głębi ostrości. Daje się również zauważyć, że obserwatorzy zdecydowanie preferują sytuację w której mają, naśladującą rzeczywistość, możliwość przeniesienia płaszczyzny ostrzenia poprzez spojrzenie na wybrany przez siebie obiekt. Preferowany jest umiarkowany, średni stopień rozmycia, choć dla dokładniejszego zamodelowania tej zależności należałoby przeprowadzić bardziej wnikliwe badanie.

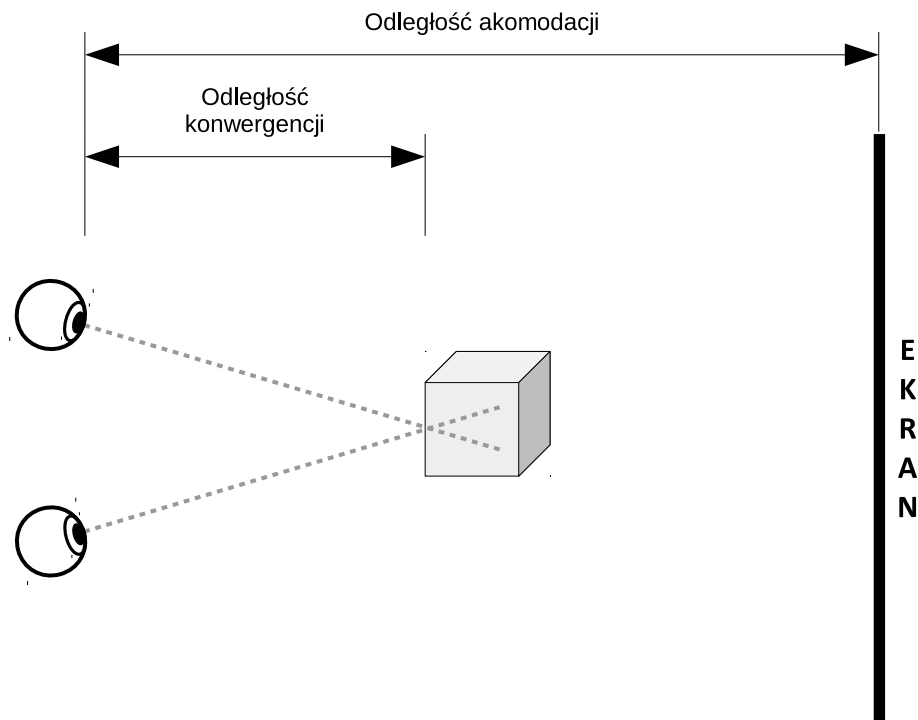
11.2.2 Łagodzenie konfliktu konwergencji-akomodacji

Interaktywna symulacja głębi ostrości może mieć na celu nie tylko wzmocnienie poczucia zanurzenia użytkownika lub poprawę wrażeń estetycznych. Jako próba naśladowania rzeczywistego zachowania aparatu widzenia, może znaleźć zastosowanie w celu uzupełnienia brakujących wskazówek o głębi podczas wyświetlania obrazu przy symulacji innych z nich, a tym samym prowadzić do redukcji zmęczenia obserwatora, które wynika z konfliktów związanych ze wspomnianymi brakami.

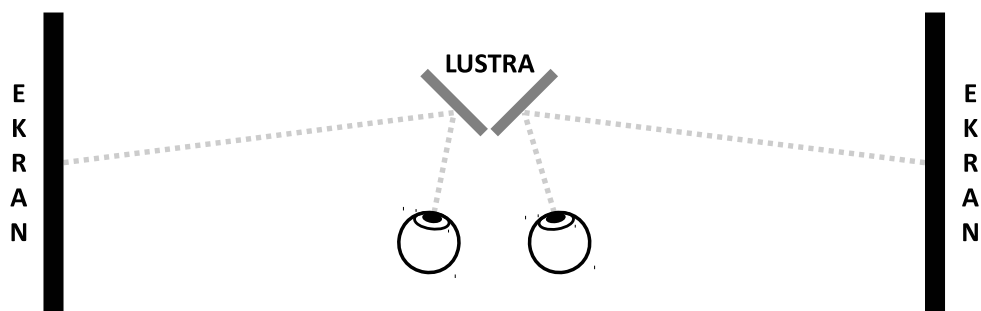
Wyświetlanie obrazu stereoskopowego związane jest z zakłóceniem naturalnej zależności pomiędzy konwergencją i akomodacją. Poprzez wyświetlanie obrazów o niezerowym rozsunięciu (ang. *disparity*) na jednej płaszczyźnie, a więc dla stałej fizycznej odległości od obserwatora, dochodzi do skonfliktowania wskazówek o głębi. Informacja o rozsunięciu zmusza aparat widzenia do reakcji konwergencyjnej w celu jego skorygowania, jednak w tym samym czasie nie występuje rozmycie wynikające z głębi ostrości, które powinno towarzyszyć temu rozsunięciu lub rozmycie to odpowiada akomodacji do zupełnie innej płaszczyzny, płaszczyzny ekranu (zob. Rys. 14).

W powiązanej z rozprawą pracy [Duc14] postawiono hipotezę mówiącą, że symulacja rozmycia wynikającego z głębi ostrości, która będzie towarzyszyła wyświetlaniu obrazu stereoskopowego, zredukuje dyskomfort powiązany z konfliktem konwergencji i akomodacji. Rozmycie na prezentowanym obrazie, które byłoby powiązane z faktyczną konwergencją oczu użytkownika, miałoby więc spełnić w postrzeganiu głębi tę samą rolę, co naturalne rozmycie obrazu rzutowanego na siatkówkę.

Przeprowadzono eksperyment z wykorzystaniem stereoskopu oraz binokularnego okulografu, który służył określaniu odległości do której dostosowana jest konwergencja oczu badanej osoby (zob. Rys. 15). Na podstawie uzyskanych wyników stwierdzono, że udało się potwierdzić wspomnianą hipotezę. Uczestnicy badania oceniali swoje zmęczenie jako mniejsze, gdy dokonywano rozmycia elementów obrazu znajdujących się bliżej lub dalej od odległości wynikającej z konwergencji oczu pod-



Rysunek 14: Problem konfliktu konwergencji i akomodacji. Przykładowy obiekt (sześcian) w wyniku zastosowania wyświetlania stereoskopowego, postrzegany jest jako znajdujący się bliżej obserwatora niż płaszczyzna ekranu. Orientacja gałek ocznych ustawiona jest tak, by doprowadzić do fuzji (połączenia) obrazów obiektu, które są rozsunięte tak, by odpowiadać innej odległości, niż odległość ekranu. Akomodacja natomiast następuje zawsze do rzeczywistej odległości ekranu.



Rysunek 15: Stereoskop wykorzystany w eksperymencie badającym wpływ interaktywnej symulacji głębi ostrości na złagodzenie skutków konfliktu konwergencji i akomodacji. Obraz dla każdego oka prezentowany jest na osobnym ekranie. Tor optyczny jest kierowany na odpowiedni ekran z użyciem luster.

czas prezentacji obrazu stereoskopowego. Podobną zależność wykazano również w zakresie występowania bólu głowy.

11.3 Głębia w obrazach o poszerzonym zakresie dynamiki luminancji

Podczas prac nad rozprawą postawiono hipotezę, że kontrast, a co za tym idzie luminancja, mogą wprowadzać mierzalną iluzję przestrzenności i współdziałać z innymi wskazówkami głębi lub modyfikować wrażenia powstałe w ich wyniku. Treść hipotezy motywowana jest często obserwowanym zjawiskiem, gdy osoby pierwszy raz widzące obraz prezentowany na ekranie *HDR*, mylą go z wyświetlaczem przedstawiającym obraz trójwymiarowy. Równie często pojawia się porównanie tak wyświetlanego obrazu ze spoglądaniem na rzeczywistą scenę przez okno [RHM11].

Problem użycia kontrastu jako narzędzia do indukowania wrażenia głębi u użytkownika komputerowego wyświetlacza został poruszony w pracy zespołu Ichihary [IKA07]. Wykorzystując standardowe monitory komputerowe i proste bodźce, udało się zaobserwować wyraźną zależność między kontrastem, a postrzeganą głębią. Powiązanie to zostało wytłumaczone związkiem ze zjawiskiem perspektywy powietrznej. Obserwacje zespołu Ichihary zostały potwierdzone dla wyświetlaczy *HDR* przez zespół Rempela [RHM11]. Z kolei Easa w pracy [EML13] wykazał, że przy wyświetlaniu bodźców na ekranie o poszerzonym zakresie dynamiki luminancji, kontrast jest jedną z najsilniejszych monokularnych wskazówek o głębi pozwalających na intuicyjne szeregowanie obiektów według odległości, z mocą sugestii porównywalną do różnicy wielkości obiektów związaną z perspektywą.

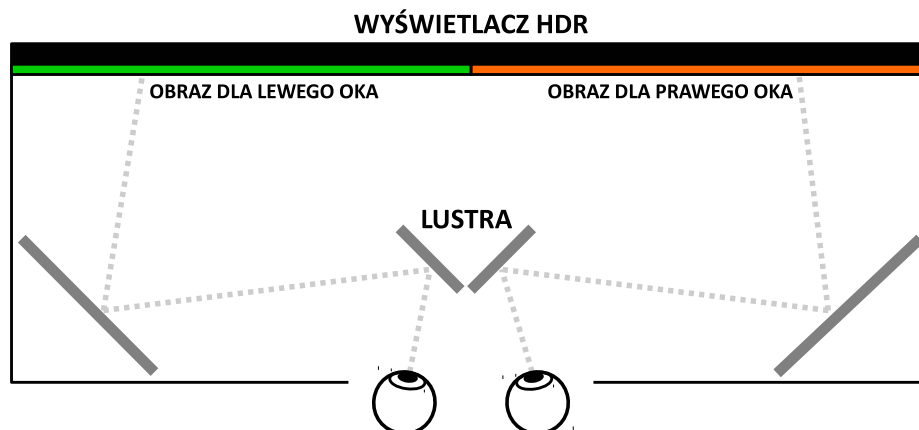
11.3.1 Badanie zależności kontrastu i wrażenia głębi

W wymienionych pracach badano jednak kontrast jako wyizolowaną wskazówkę głębi. Przedmiotem opisanych w rozprawie badań jest natomiast analiza wpływu kontrastu na postrzeganie głębi w sytuacji, gdy obecne są także, uznawane za silniejsze, wskazówki binokularne w postaci obrazu stereoskopowego. W tym celu wykorzystano 47-calowy wyświetlacz *HDR*, który posłużył do budowy stereoskopu (zob. Rys. 16 i Rys. 17).

W wyniku przeprowadzonych badań udało się potwierdzić postawioną wcześniej hipotezę, jednak uzyskany efekt zdaje się więc być zbyt słabym, aby mógł być wykorzystany do wzmocnienia lub osłabienia wrażenia głębi pochodzącego ze wskazówki stereoskopowej. Ta ostatnia jest znacznie silniejsza i poza okolicą progu jej wykrywalności, jej wpływ na ocenę odległości jest przeważający. Jednak kontrast mógłby stać się częścią przekazu o przestrzenności sceny w warunkach, gdy dodatkowe wzmocnienie rozsunęcia klatek dla obojga oczu wiązałoby się z wyjściem poza strefę komfortu widzenia stereoskopowego.

11.3.2 Kontrast i głębia a poczucie realizmu

Chcąc poznać przyczynę raportowanego przez zespół Rempela silnego poczucia realności obrazów oglądanych na wyświetlaczu *HDR* [RHM11], przygotowano następny eksperyment, tym razem bazujący na obrazach rzeczywistej sceny (zob. Rys. 18).



Rysunek 16: Stereoskop wykorzystany w eksperymencie badającym wpływ kontrastu na postrzeganie głębi. Obraz dla każdego oka prezentowany jest na dwu częściach jednego ekranu. Tor optyczny jest kierowany na odpowiednią połowę ekranu z użyciem luster.



Rysunek 17: Stanowisko do badań nad kontrastem. Po lewej stronie widoczna obudowa stereoskopu opartego o wyświetlacz *HDR*. Z prawej strony obudowa kontenera z rzeczywistą sceną zawierającą fizyczne obiekty, która posłużyła do uzyskania fotografii dla badania opisanego w rozprawie.



Rysunek 18: Scena zbudowana na potrzeby przeprowadzenia eksperymentu. Przedstawiony obraz stanowi pokrycie pełnego pola widzenia dla jednego oka przy obrazowaniu stereoskopowym. Pionowy charakter klatki ma związek ze sposobem działania użytego stereoskopu, gdzie horyzontalny, panoramiczny ekran został podzielony na dwie części: lewą i prawą, przedstawiające klatki dla poszczególnych oczu.

Postawiono hipotezę o wpływie poszerzonego zakresu dynamiki luminancji podczas wizualizacji obrazów pochodzących ze świata rzeczywistego, na wywołanie u oglądającego wrażenia realizmu, prowadzącego do postrzegania sceny jako rzeczywistych obiektów, a nie ich obrazów na płaszczyźnie. Celem przeprowadzonego badania stała się więc próba pomiaru zmiany odczucia realizmu, zależnie od kontrastu i binokularnych wskazówek o głębi.

Ważnym spostrzeżeniem wynikającym z analizy wyników tego eksperymentu jest występowanie w populacji charakterystycznych grup cechujących się silnymi preferencjami odnośnie badanych czynników. Choć przebadana próba jest zbyt mała by próbować oszacować licznosc wspomnianych grup w populacji, liczba osób które przejawiały wspólne preferencje we wspomnianych wyżej grupach jest stosunkowo duża. Uzasadnione więc wydaje się wnioskowanie, że każda z nich reprezentuje istotną liczbę osób stanowiących część populacji.

W kwestii preferencji dotyczących kontrastu, można zauważyć trend prowadzący do częstszego wyboru wzmocnionego, niż osłabionego kontrastu jako tego bardziej realnego. Takie zjawisko zostało opisane już wcześniej [YMMS06], gdzie badani wy-

bierali mocniejszy od neutralnego kontrast nawet w przypadku gdy mogli porównać prezentowany obraz z rzeczywistą referencją. Można próbować tłumaczyć tę tendencję wyczuleniem na charakterystyczne cechy wyświetlacza *HDR*, które były dostrzegane przez badanych jako nowe, wyjątkowe elementy. Być może wzmocniony kontrast był odbierany jako bardziej realny, gdyż kompensował braki w innych aspektach realizmu prezentowanego obrazu.

11.4 Podsumowanie

Powstawanie wrażenia głębi sceny oraz powiązanego z nim poczucia realizmu oglądanego, syntezywanego komputerowo obrazu, obejmuje wiele elementów ludzkiego systemu postrzegania. Elementy te w naturalnych warunkach pobudzane są szerokim spektrum bodźców, spośród których wiele nie pojawia się przy obserwowaniu wizualizacji wirtualnej sceny. Na łamach rozprawy dokonano przeglądu różnych aspektów tego problemu, koncentrując się na zaproponowaniu technik pozwalających zwiększyć ilość informacji przestrzennej zawartej w dostarczanych odbiorcy obrazach. Techniki te wykorzystują mechanizm widzenia stereoskopowego, wrażliwość aparatu widzenia na kontrast, pomiar i wykorzystanie chwilowego kierunku spojrzenia w czasie zbliżonym do rzeczywistego i wynikające z akomodacji rozmycie obrazu.

Publikacje własne

- [Baz10] Bartosz Bazyluk. Eye tracking in virtual environments: The study of possibilities and the implementation of gaze-point dependent depth of field. In *Proceedings of Central European Seminar on Computer Graphics 2010*, Vienna, Austria, 2010. VUT.
- [BM14a] Bartosz Bazyluk and Radosław Mantiuk. Gaze-driven object tracking based on optical flow estimation. *Lecture Notes in Computer Science (Proc. of ICCVG 2014)*, 8671, 2014.
- [BM14b] Bartosz Bazyluk and Radosław Mantiuk. Rozszerzenie zakresu interakcji z grą komputerową z wykorzystaniem danych okulograficznych. In Sambor Gucza, editor, *Widziane inaczej. Z polskich badań eyetrackingowych*, volume 20 of *Studia Naukowe*, pages 121–135. Wydawnictwo Naukowe Instytutu Kulturologii i Lingwistyki Antropocentrycznej, Warszawa, 2014.
- [Duc14] Duchowski, Andrew T and House, Donald H and Gestring, Jordan and Wang, Rui I and Krejtz, Krzysztof and Krejtz, Izabela and Mantiuk, Radosław and Bazyluk, Bartosz. Reducing visual discomfort of 3D stereoscopic displays with gaze-contingent depth-of-field. In *Proceedings of the ACM Symposium on Applied Perception*, pages 39–46. ACM, 2014.
- [MB14] Radosław Mantiuk and Bartosz Bazyluk. Dokładność pomiaru kierunku patrzenia. In Sambor Gucza, editor, *Widziane inaczej. Z polskich badań eyetrackingowych*, volume 20 of *Studia Naukowe*, pages 9–18. Wydawnictwo Naukowe Instytutu Kulturologii i Lingwistyki Antropocentrycznej, Warszawa, 2014.
- [MBM13] Radosław Mantiuk, Bartosz Bazyluk, and Rafał K. Mantiuk. Gaze-driven object tracking for real time rendering. *Computer Graphics Forum*, 32(2):163–173, 2013.
- [MBT11] Radosław Mantiuk, Bartosz Bazyluk, and Anna Tomaszewska. Gaze-dependent depth-of-field effect rendering in virtual environments. In *Serious Games Development and Applications*, pages 1–12. Springer, 2011.
- [MCC⁺11] R Mantiuk, J Cybularczyk, M Cichowicz, M Smyk, and B Bazyluk. Hdr photographic pipeline for camera modules in mobile devices. In *Eurographics 2011 - Areas Papers*, pages 77–78. The Eurographics Association, 2011.
- [MKNB12] Radosław Mantiuk, Michał Kowalik, Adam Nowosielski, and Bartosz Bazyluk. Do-it-yourself eye tracker: Low-cost pupil-based eye tracker for computer graphics applications. *Lecture Notes in Computer Science (Proc. of MMM 2012)*, 7131:115–125, 2012.

- [Pet14] Peter Vangorp and Rafał K. Mantiuk and Bartosz Bazyluk and Karol Myszkowski and Radosław Mantiuk and Simon J. Watt and Hans-Peter Seidel. Depth from HDR: Depth induction or increased realism? In *Proceedings of the ACM Symposium on Applied Perception*, pages 71–78. ACM, 2014.
- [TB10] Anna Tomaszewska and Bartosz Bazyluk. Effective real-time computer graphics processing based on depth-of-field effect. *Pomiary Automatyka Kontrola*, 7:675–677, 2010.

Wybrane pozycje literaturowe

- [CF10] Alan Chalmers and Andrej Ferko. Levels of realism: From virtual reality to real virtuality. In *Proceedings of the 24th Spring Conference on Computer Graphics, SCCG '08*, pages 19–25, New York, NY, USA, 2010. ACM.
- [DPHW11] Andrew T. Duchowski, Brandon Pelfrey, Donald H. House, and Rui Wang. Measuring gaze depth with an eye tracker during stereoscopic display. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, APGV '11*, pages 15–22, New York, NY, USA, 2011. ACM.
- [EML13] Haider K. Easa, Rafał K. Mantiuk, and Ik Soo Lim. Evaluation of monocular depth cues on a high-dynamic-range display for visualisation. *ACM Trans. Appl. Percept.*, 10(3):16, 2013.
- [Fer03] James A. Ferwerda. Three varieties of realism in computer graphics. In *SPIE Human Vision and Electronic Imaging 2003*, pages 290–297, Bellingham, 2003. SPIE.
- [HNA⁺11] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer. *Eye Tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.
- [IKA07] Shigeru Ichihara, Norimichi Kitagawa, and Hiromi Akutsu. Contrast and depth perception: Effects of texture contrast and area contrast. *Perception*, 36:686–695, 2007.
- [KvI⁺16] Jari Kangas, Oleg Špakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. Feedback for smooth pursuit gaze tracking based control. In *Proceedings of the 7th Augmented Human International Conference 2016, AH '16*, pages 6:1–6:8, New York, NY, USA, 2016. ACM.
- [LK81] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. *Proc. of the 7th International Joint Conference on Artificial Intelligence*, 2:674–679, 1981.
- [PKS⁺17] Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A. Cooper, and Gordon Wetzstein. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences*, 114(9):2183–2188, 2017.
- [PSD80] Michael I Posner, Charles R Snyder, and Brian J Davidson. Attention and the detection of signals. *Journal of experimental psychology: General*, 109(2):160–174, 1980.

- [RFN92] Jyrki Rovamo, Rauli Franssila, and Risto Näsänen. Contrast sensitivity as a function of spatial frequency, viewing distance and eccentricity with and without spatial noise. *Vision research*, 32(4):631–637, 1992.
- [RHM11] Allan G. Rempel, Wolfgang Heidrich, and Rafał Mantiuk. The role of contrast in the perceived depth of monocular imagery. Technical Report TR-2011-07, University of British Columbia, 2011.
- [SG00] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.
- [tob11] Accuracy and precision test method for remote eyetrackers. Technical report, Tobii Technology, 2011.
- [VBG13] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, September 2013.
- [YMMS06] Akiko Yoshida, Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. Analysis of Reproducing Real-World Appearance on Displays of Varying Dynamic Range. *Computer Graphics Forum (Proc. of Eurographics)*, 25(3):415–426, September 2006.
- [ZJM09] Philip G. Zimbardo, Robert L. Johnson, and Vivian McCann. *Psychologia. Kluczowe Koncepcje*, volume 3. Wydawnictwo Naukowe PWN, 6 edition, 2009.

2017-10-13
Bartosz Bazyli

Warszawa, 24 lipca 2017 r.

Prof. dr hab. inż. Przemysław Rokita
Kierownik Zakładu Grafiki Komputerowej
Instytut Informatyki Politechniki Warszawskiej
Nowowiejska 15/19, 00-665 Warszawa

**RECENZJA ROZPRAWY DOKTORSKIEJ
DLA RADY WYDZIAŁU INFORMATYKI ZACHODNIOPOMORSKIEGO
UNIwersYTETU TECHNOLOGICZNEGO W SZCZECINIE**

Tytuł rozprawy:

**Metody modelowania percepcji głębi oraz widzenia kierunkowego
na potrzeby obrazowania komputerowego**

Autor rozprawy:

mgr inż. Bartosz Bazyluk

*1. Jakie zagadnienie naukowe jest rozpatrzone w pracy (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, inny)?**

Praca pt. „Metody modelowania percepcji głębi oraz widzenia kierunkowego na potrzeby obrazowania komputerowego” liczy 155 stron tekstu i zawiera 165 pozycji bibliografii. Autor przedstawia w niej opracowane przez siebie, oraz w zespołach których był członkiem, algorytmy i ich badania dla potrzeb uwzględnienia w komputerowej syntezie obrazów, widzenia kierunkowego oraz percepcji głębi. Praca składa się z 3 rozdziałów, wprowadzenia, podsumowania oraz bibliografii.

We wprowadzeniu zawarto definicję problemu, tezę, cel pracy oraz przedstawiono organizację treści rozprawy. Jako podstawowy cel pracy określono „zbadanie wybranych aspektów sposobu widzenia człowieka w kontekście ich użyteczności dla komputerowych metod wizualizacji głębi obrazu, a następnie zaproponowanie i sprawdzenie skuteczności rozwiązań umożliwiających wykorzystanie pozyskanej wiedzy w procesie syntezy oraz prezentacji grafiki komputerowej”.

W rozdziale 1. Autor przedstawił wprowadzenie do problematyki percepcji realizmu i głębi obrazów, uwagi wzrokowej oraz kierunkowości widzenia. Opisał w nim elementy wpływające na poczucie realizmu w symulacji obrazowej, ze szczególnym uwzględnieniem postrzegania przestrzenności sceny i kierunkowości widzenia obserwatora.

W rozdziale 2. Doktorant przedstawił problematykę modelowania i pomiarów kierunkowości widzenia. Opisał techniczne możliwości oraz problemy związane ze śledzeniem kierunku spojrzenia za pomocą okulografii. Najistotniejszym fragmentem tej części rozprawy są podrozdziały 2.3-2.4, w których Autor przedstawił opracowany w ramach pracy, oryginalny algorytm identyfikacji

syntetycznych obiektów śledzonych wzrokiem przez obserwatora oraz adaptację tego algorytmu dla potrzeb analizy materiałów wideo.

Rozdział 3 zawiera opis implementacji oraz przeprowadzonych badań symulacji głębi ostrości w grafice komputerowej, wykorzystującej dane okulograficzne do wyznaczenia kierunku spojrzenia obserwatora. W rozdziale tym przedstawiono również badania wpływu kontrastu i głębi na poczucie realizmu, przeprowadzone na stereoskopowych wyświetlaczach o poszerzonym zakresie dynamiki luminancji, w których doktorant brał udział. Najistotniejszym fragmentem tej części pracy jest podrozdział 3.1 zawierający opis oryginalnej implementacji i badań symulacji głębi ostrości z uwzględnieniem kierunku spojrzenia obserwatora.

W podsumowaniu Autor przedstawił syntetyczne wnioski końcowe oraz zarys możliwych dalszych prac.

Praca ma charakter zarówno teoretyczny jak i doświadczalny. Autor na podstawie znajomości rozległej teorii percepcji wzrokowej oraz własnych doświadczeń implementacji algorytmów grafiki komputerowej, opracował oryginalne rozwiązania algorytmiczne i implementacyjne, jak też przeprowadził ich badania prowadzące do ważnych wniosków praktycznych.

2. *Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle) świadczący o dostatecznej wiedzy autora. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący? **

Analiza źródeł przedstawiona w pracy jest bardzo dobra. Cały tekst pracy świadczy o znajomości przez Doktoranta obszernej literatury z dziedziny percepcji wzrokowej, okulografii oraz grafiki komputerowej. Opracowane przez niego rozwiązania teoretyczne oraz zaimplementowane algorytmy i ich testy, będące praktyczną weryfikacją zaproponowanych rozwiązań, świadczą o bardzo dobrej umiejętności analize literatury, poprawnego formułowania na jej podstawie wniosków oraz opracowywania zaawansowanych rozwiązań programistycznych.

3. *Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione? **

Cel pracy sformułowany w jej wstępie, tzn. zaproponowanie i przebadanie rozwiązań umożliwiających wykorzystanie wiedzy o kierunkowości i przestrzenności widzenia w procesie syntezy oraz prezentacji grafiki komputerowej został osiągnięty. Autor opracował nowe rozwiązania wykorzystujące w analizie obrazów i grafice komputerowej dane okulograficzne. Rozwiązania te mają istotne zalety w stosunku do wcześniejszych znanych z literatury i jako takie stanowią istotny wkład w rozwój algorytmów wizualizacji i analizy obrazów. Rozwiązania te Autor opracował w oparciu o własne doświadczenia, w tym uzyskane w ramach prac zespołów badawczych oraz analizę rozwiązań opisanych w literaturze. Przyjęta metodologia, założenia i uproszczenia są zgodne z obecnie stosowanymi w grafice komputerowej i analizie obrazów.

4. *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?* *

W rozprawie przedstawiono bardzo obszerny materiał stanowiący połączenie wyników indywidualnych doświadczeń i przemyśleń Doktoranta oraz prac w ramach zespołów badawczych, których był członkiem. Autor dokonał kompleksowej analizy wybranych problemów modelowania widzenia kierunkowego oraz percepcji głębi na potrzeby grafiki komputerowej. Efektem tych prac było sformułowanie własnych, nowych propozycji rozwiązań. Do najistotniejszych elementów dorobku Doktoranta należy w szczególności:

- oryginalna probabilistyczna metoda identyfikacji przedmiotu uwagi w scenie wizualizowanej za pomocą grafiki komputerowej, bazująca na danych okulograficznych oraz analizie położenia i wektorów prędkości punktu skupienia wzroku oraz wyodrębnionych elementów syntetyzowanego obrazu (rozdz. 2.3),
- oryginalna probabilistyczna metoda identyfikacji przedmiotu uwagi w sekwencjach wideo bazująca na danych okulograficznych oraz analizie przepływu optycznego w sekwencji filmowej (rozdz. 2.4.1),
- implementacja i przebadanie systemu wirtualnej rzeczywistości symulującego zjawisko głębi ostrości na bazie danych z okulografu (rozdz. 3.1).

Rezultaty prac uzyskane przez Doktoranta były prezentowane na renomowanych międzynarodowych konferencjach naukowych (m.in. *ACM Symposium on Applied Perception*, *International Conference on Computer Vision and Graphics*, *International Conference on Multimedia Modeling*) oraz w czasopiśmie naukowym, w tym z „listy filadelfijskiej” (doktorant był współautorem pracy opublikowanej w *Computer Graphics Forum*).

5. *Czy autor wykazał umiejętność poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?* *

Praca została starannie zredagowana. Autor wykazał w rozprawie umiejętność poprawnego i przekonującego przedstawienia swoich przemyśleń, wniosków oraz propozycji rozwiązań. Wykazał się też umiejętnością zwięzłej prezentacji wyników swych prac. Praca stanowi dobre wprowadzenie do problematyki badań w obrębie problemów modelowania percepcji głębi oraz widzenia kierunkowego na potrzeby grafiki komputerowej.

6. *Jakie są słabe strony rozprawy i jej główne wady?* *

W pracy nie znalazłem błędów merytorycznych natomiast ma ona kilka istotnych wad oraz mankamentów od strony warsztatowej. Należą do nich w szczególności:

- brak informacji o kosztach obliczeniowych i wymaganiach sprzętowych zaproponowanych algorytmów oraz dyskusji ich wpływu na opóźnienia generowane w toku przetwarzania i generowania klatek,

- brak syntetycznej (liczbowej) specyfikacji oraz dyskusji błędów pomiarowych uzyskiwanych w trakcie analizy danych okulograficznych i zestawienia ich z wymaganiami konkretnych algorytmów syntezy obrazów, w których zostały wykorzystywane,
- w rozdz. 3.1.3.1 na str.111 w opisie eksperymentu przyjęto arbitralnie stopnie rozmycia bez wyjaśnienia ich powiązania z parametrami gałki ocznej, nie podano przyjętych parametrów symulowanego układu optycznego, tylko liczby bez jednostek (dla symulacji zjawisk optycznych zachodzących w gałce ocznej należałoby podawać średnicę źrenicy w milimetrach, a nie w jednostkach względnych),
- praca przedstawia zaawansowane zagadnienia realistycznej grafiki komputerowej, a w szczególności przetwarzania i generowania sekwencji obrazów, natomiast jedyne udostępnione ilustracje, to niskiej jakości obrazy załączone (wydrukowane) w treści pracy – istotną wadę stanowi brak udostępnienia przez Doktoranta (np. na płycie CD/DVD), obrazów wysokiej (źródłowej) jakości oraz materiałów wideo prezentujących testy (np. opisane na str. 66, 74, 84, 104) oraz wyniki (zwłaszcza te uzyskane w rozdz. 2.4.3 oraz 3.1),
- tekst pracy sugeruje (np. na str. 94), że jej oryginalnym wkładem jest opracowanie koncepcji interaktywnej symulacji głębi ostrości – o ile implementacja tej symulacji i jej przebadanie są rzeczywiście oryginalnym wkładem Doktoranta, to sama koncepcja była już opisana w literaturze wiele lat temu np. w artykule wymienionym w bibliografii pracy: *Przemysław Rokita „Generating depth of-field effects in virtual reality applications” IEEE Computer Graphics and Applications, vol.16(2), pp.18–21, March 1996,*
- w tekście często zamiast konkretnych danych liczbowych można znaleźć b.nieprecyzyjne lub abstrakcyjne określenia – np. na str.5 „czas rzeczywisty”, str.37 „wystarczający do precyzyjnego określenia”, str.40 „zmierzenie kierunku patrzenia z idealną dokładnością”, str.42 „doskonałej precyzji”, str.60 „odpowiednio blisko”,
- w rozdz. 2.1.4.2 na str. 34-35 brak zadowalającego wyjaśnienia źródła omawianych błędów - nie rozważono w szczególności wpływu na ich powstawanie i minimalizację, położenia kamery okulografu w stosunku do oka (lewo/prawo, góra/dół),
- w rozdz.2.1.5.3 na str. 39 nie dostrzeżono możliwości wykorzystania tanich czujników inercyjnych MEMS do śledzenia lub wspomagania śledzenia ruchów głowy,
- w całej pracy, a w szczególności w rozdz. 2.4.3 w dyskusji zastosowań algorytmu opracowanego dla potrzeb interfejsów użytkownika, Doktorant nie dostrzegł możliwości wykorzystania opracowanych rozwiązań dla potrzeb osób niepełnosprawnych – jest to moim zdaniem bardzo istotne pole zastosowań okulografii i ew. dalszych prac Autora,
- na wielu ilustracjach (np. na rysunkach 2.18, 2.20, 2.22, 2.23, 2.27, 2.29, 3.12, 3.23) opisy tekstowe są w języku angielskim – praca napisana jest w języku polskim więc opisy ilustracji też powinny zostać przetłumaczone,
- na str. 88 w tekście jest odwołanie do ilustracji (rys.5), której brak w tekście.

Podsumowując tą dość długą listę uwag krytycznych chciałbym zauważyć, że wymienione wady mają w większości charakter redakcyjny i nie podważają wartości merytorycznej dorobku Doktoranta.

7. *Jaka jest praktyczna przydatność rozprawy?* *

Praktyczna przydatność rozprawy jest potencjalnie duża. Główny wątek pracy - optymalizacja algorytmów modelowania percepcji głębi oraz widzenia kierunkowego dla potrzeb grafiki komputerowej ma duże znaczenie praktyczne. Opracowane rozwiązania mogą zostać wykorzystane m.in. w różnego rodzaju systemach symulacji wizualnej, wirtualnej rzeczywistości oraz grach komputerowych. Ponadto opracowane algorytmy mogą zostać wykorzystane do tworzenia efektywniejszych graficznych interfejsów użytkownika, jak też percepcyjnej analizy sekwencji wideo.

8. *Do której z następujących kategorii Recenzent zalicza rozprawę:*

a/ nie spełniająca wymagań stawianych rozprawom doktorskim przez obowiązujące przepisy

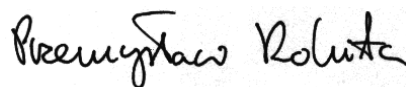
b/ wymagająca wprowadzenia poprawek i ponownego recenzowania

c/ spełniająca wymagania

d/ spełniająca wymagania z wyraźnym nadmiarem

e/ wybitnie dobra, zasługująca na wyróżnienie

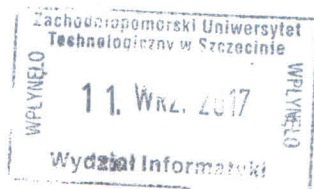
Podsumowując stwierdzam, że recenzowana praca zawiera prezentację istotnego dorobku naukowego Doktoranta. Dorobek ten poza samą rozprawą został zaprezentowany i poddany weryfikacji w ramach prezentacji na renomowanych międzynarodowych konferencjach oraz publikacji w czasopiśmie naukowych – w tym z „*listy filadelfijskiej*”. Uważam, że recenzowana praca spełnia z wyraźnym nadmiarem wymagania stawiane rozprawom doktorskim, określone w art. 13 ust. 1 obowiązującej ustawy z dnia 14 marca 2003r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (z późniejszymi zmianami). W związku z tym wnoszę o dopuszczenie jej Autora do publicznej obrony.



.....
podpis

* Powyższe pytania mają charakter pomocniczy. Recenzja jest tak sformułowana, by można ją było odczytywać bez treści pytań.

dr hab. inż. Adam Wojciechowski
Instytut Informatyki
Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej
Politechniki Łódzkiej



Łódź, 08.08.2017

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy: Metody modelowania percepcji głębi oraz widzenia kierunkowego na potrzeby obrazowania komputerowego

Autor rozprawy: mgr inż. Bartosz Bazyluk

Promotor: dr hab. inż. Radosław Mantiuk, prof. ZUT

Dziedzina: Nauki techniczne

Dyscyplina: Informatyka

1. Jaki jest problem naukowy (teza) rozprawy oraz czy został on trafnie i jasno sformułowany?

Problem stymulowania ludzkiego systemu postrzegania w sposób pozwalający na wywołanie wrażeń zbliżonych do tych, których doświadcza człowiek, obcując ze światem rzeczywistym, jest złożonym zagadnieniem naukowym. W zakresie graficznych systemów komputerowych zagadnienie można zdekomponować na szereg składowych problemów związanych z syntezą i wyświetlaniem cyfrowych obrazów środowisk wirtualnych. Jednym z głównych aspektów decydujących o poziomie foto-realizmu obserwowanego obrazu jest wrażenie jego przestrzenności determinowane przez tzw. wskaźniki głębi, które można podzielić na jednooczne i dwuoczne. Autor właściwie przeanalizował i zidentyfikował deficyty w zakresie wskaźników głębi syntetycznych obrazów cyfrowych. Umotywował swoją analizę problemu dogłębnym przeglądem literatury, który dostarczył mu szeregu inspiracji na rzecz zaproponowanych eksperymentów naukowych.

Autor postanowił zbadać jakość postrzegania głębi obrazu i powiązanego z nią foto-realizmu obrazu cyfrowego dla opracowanych przez siebie technik i metod kompensacji deficytów naturalnych mechanizmów towarzyszących postrzeganiu rzeczywistości przez system wzrokowy człowieka, w tym m.in.: dynamicznego, interaktywnego zarządzania głębią ostrości, powiązanego z akomodacją i konwergencją, czy wpływu kontrastu obrazów o wysokiej dynamice luminancji na postrzeganie głębi obrazu stereoskopowego.

Jako główne narzędzie niezbędne do przeprowadzenia badań Autor wskazał efektywny mechanizm estymacji kierunku patrzenia, który pozwala na śledzenie ruchów gałek ocznych obserwatora w czasie rzeczywistym i na tej podstawie szacuje podmiot koncentracji wzroku (zainteresowania) widza w obrazie. Z racji na duży koszt profesjonalnych narzędzi do śledzenia ruchów gałek ocznych (tzw. *eye tracker*) Autor podjął próbę stworzenia własnego, niskobudżetowego, narzędzia do wideookulografii, które, jak wykazał w eksperymentach, cechuje się porównywalną do urządzenia profesjonalnego (*SensoMotoric Instruments RED250*) jakością pozyskiwanych danych pomiarowych.

W kontekście badań naukowych Autor sformułował tezę rozprawy, brzmiącą następująco:

„Wykorzystanie modeli mechanizmu percepcji głębi człowieka umożliwia poprawę jakości wizualizacji obrazów trójwymiarowych. Dotyczy to takich mechanizmów jak: kierunkowość widzenia, czułość na kontrast w szerokim zakresie dynamiki luminancji, mechanizm akomodacji, widzenie stereoskopowe oraz subiektywne wrażenie realizmu obrazu.”

Głównym celem pracy było zaś *„zbadać wybranych aspektów sposobu widzenia człowieka w kontekście ich użyteczności dla komputerowych metod wizualizacji głębi obrazu, a następnie zaproponowanie i sprawdzenie skuteczności rozwiązań umożliwiających wykorzystanie pozyskanej wiedzy w procesie syntezy oraz prezentacji grafiki komputerowej”*.

Zarówno cel jak i teza rozprawy są sformułowane właściwie. Pewnego wyjaśnienia jedynie może wymagać fragment tezy, z którego wynika, że „subiektywne wrażenie realizmu obrazu” zostało zakwalifikowane do „mechanizmów percepcji głębi obrazu”.

2. Czy tematyka rozprawy jest aktualna lub dostatecznie ważna?

Tematyka rozprawy jest ważna i aktualna, szczególnie w odniesieniu do zastosowań w obszarze dynamicznie rozwijającego się przemysłu rozrywkowego (filmy, gry wideo, itp.) oraz możliwych zastosowań w obszarze szeroko pojętej interakcji człowiek-komputer, w tym dla wizualizacji i testowania środowisk wirtualnych, czy materiałów filmowych (badania marketingowe, interfejsy dedykowane osobom niepełnosprawnym, itp.). Osiągnięcia pracy wpisują się w proces lepszego zrozumienia oraz modelowania mechanizmów postrzegania i syntezy obrazów cyfrowych, tak aby był on możliwie bliski procesowi percepcji środowiska rzeczywistego.

Zastosowania osiągnięć badawczych doktoranta zostały zasygnalizowane i przedstawione w dysertacji. Zaproponowane przez Autora sposoby użycia metody szacowania kierunku postrzegania sceny bardzo dobrze nakreślają perspektywę współczesnych badań w rozważanym obszarze. W pracy wymienione zostały przykładowo: interaktywne zarządzanie głębią ostrości syntezy obrazu, które ma istotne znaczenie przy poprawie zanurzenia obserwatora w środowiskach wirtualnych, optymalizacja niezbędnych obliczeń poza obszarem skupienia obserwatora, służących usprawnieniu komputerowej syntezy obrazu, czy też wykorzystanie mechanizmów postrzegania elementów interfejsu graficznego do jego zdalnej obsługi.

Autor nawiązał również do bardzo ciekawych zagadnień związanych z uwagą wzrokową (np. dysonans spojrzenia i uwagi), która znajduje się w centrum zainteresowania badań marketingowych, badań nad bezpieczeństwem publicznym, czy zastosowań medycznych. W pracy zasygnalizowane zostały również kwestie istotności wskaźników głębi, które mają przełożenie na dalszy rozwój procesu modelowania percepcji głębi. Ciekawym rozwinięciem pracy i naturalną kontynuacją badań Autora mogłoby być zgłębienie powiązań kognitywnej uwagi wzrokowej i fizjologicznych aspektów widzenia (np. koherencja mapy uwagi (ang. *saliency map*) i towarzyszących oglądanej scenie reakcji fizjologicznych widza).

3. Czy autor rozwiązał postawiony problem i użył do tego właściwych metod?

Praca ma charakter eksperymentalno-analityczny i obejmuje systematyczny wywód naukowy zmierzający do udowodnienia postawionej tezy, poprzez realizację cząstkowych eksperymentów i weryfikowanie towarzyszących im hipotez roboczych. Udowodnienie postawionej tezy wymagało realizacji celów szczegółowych, wśród których Autor jawnie wymienia:

- zbadanie metod filtracji danych okulograficznych na potrzeby identyfikacji elementu, na którym spoczywa uwaga obserwatora w środowiskach trójwymiarowych i zaproponowanie metody dedykowanej dynamicznym wizualizacjom wirtualnych środowisk;
- zbadanie możliwych zastosowań wiedzy o punkcie skupienia wzroku w wizualizacji wirtualnego środowiska zarówno w kontekście przyspieszenia renderowania jak i elementów interfejsu użytkownika;
- opracowanie metody interaktywnej, opartej na danych okulograficznych, symulacji zjawiska głębi ostrości, zbadanie jej wpływu na odbiór wizualizacji wirtualnego środowiska przez użytkownika oraz zbadanie możliwości zredukowania za jej pomocą dyskomfortu konwergencji i akomodacji w wyświetlaczach stereoskopowych;
- zbadanie wpływu kontrastu w stereoskopowych wyświetlaczach o poszerzonym zakresie dynamiki luminancji na postrzeganie głębi oraz wrażenie realności oglądanego obrazu;

W zakresie identyfikacji elementu, na którym skupia się wzrok obserwatora, Autor przeprowadził wnikliwą analizę technik okulograficznych i zaproponował **własny system wideookulograficzny DIY**.

Autorskie rozwiązanie stosuje światło podczerwone do usprawnienia segmentacji źrenicy oka i na podstawie geometrycznych relacji położenia środka źrenicy, ekranu i kamery system szacuje punkt koncentracji wzroku. Model matematyczny, używany do mapowania danych okulograficznych na współrzędne ekranu, został dobrany eksperymentalnie, zaś jego współczynniki oblicza się na etapie kalibracji, a następnie wykorzystuje w trakcie właściwych badań pomiarowych. Aby udowodnić jakość systemu okulograficznego (precyzję, dokładność i powtarzalność wyników) Autor zaproponował **własne miary błędu** i w ramach eksperymentów badawczych wykazał porównywalną z profesjonalnym urządzeniem jakość pomiarów swojego systemu. System DIY cechował się oczywiście mniejszą częstotliwością akwizycji danych (30 Hz zamiast 250 Hz), ale w ramach wprowadzonych metryk jakościowych wyniki były porównywalne.

Eksperymenty weryfikujące jakość pomiarów wykazały konieczność filtracji danych celem, stabilnego w czasie, szacowania uwagi badanej osoby. W tym celu została opracowana **autorska metoda GDOT**, stosująca probabilistyczny model do określenia obiektu, na który z największym prawdopodobieństwem koncentruje się obserwator. W algorytmie obliczenia prawdopodobieństwa istotnym usprawnieniem wydaje się potraktowanie rozkładów błędów danych okulograficznych jako połowy wzorcowego rozkładu normalnego. Algorytm uwzględnia zarówno punkt skupienia wzroku jak również trajektorię jego zmian, co pozwala na modelowanie tzw. płynnych pościgów wzroku (ang. *smooth pursuit*).

W opisie, moim zdaniem, brakuje jednak informacji jak jest definiowana pozycja obiektu, jak jest liczona i jak na wyniki wpływa wielkość, kształt obiektów sceny. Nie znalazłem informacji co jest położeniem estymowanego obiektu i w jakiej płaszczyźnie liczona jest odległość euklidesowa pomiędzy tym punktem, a estymowanym punktem skupienia. W rozdziale 2.3.1.2 pojawia się wprowadzenie próba zdefiniowania „potencjalnych odbiorców uwagi”, jednak rysunek 2.16 nie charakteryzuje tzw. „punktowych obiektów”, w sytuacji, gdy wewnątrz sceny znajduje się kilka obiektów przestrzennych (smok, piłka) i na obraz naniesione są małe okręgi o niewyjaśnionym charakterze i genezie - ich znaczenie można wywnioskować dopiero z dalszej części pracy. Pod koniec rozdziału 2.3.1.6 Autor pisze, że „metoda GDOT zwraca informację, na którym obiekcie, a nie na którym elemencie obiektu, użytkownik skupia uwagę”. Dlaczego potencjalni odbiorcy uwagi znajdują się na płaskich ścianach, podczas gdy obserwator koncentruje się na obiektach sceny lub ich fragmentach?

Zaproponowane przez Autora dysertacji rozwiązanie zostało porównane eksperymentalnie z konkurencyjnymi, popularnymi metodami I-DT i I-VT. Opracowano w tym celu **oryginalną i ciekawą metrykę Q** (str. 65) badania jakości śledzenia obiektów sceny w oparciu o rozkład prawdopodobieństwa błędów szacowania pozycji na etapie kalibracji, uwzględniającą zarówno częstotliwość jak i czas, w którym występują błędne pomiary. Jest ona zdefiniowana właściwie i wykazuje przewagę metody GDOT nad rozwiązaniami konkurencyjnymi.

Zaproponowana metoda szacowania punktu skupienia wzroku znajduje szerokie zastosowanie w różnych zadaniach stawianych przed współczesnymi systemami komputerowymi. W rozdziale 2.4 Autor **wykazał eksperymentalnie aż trzy obszary zastosowań autorskiej metody GDOT**. Pierwsza dotyczyła śledzenia punktów charakterystycznych obrazu wideo (telewizyjna transmisja meczu piłki nożnej), druga przyspieszenia syntezy obrazu generowanego komputerowo, uwzględniającego uproszczoną symulację głębi ostrości na granicy obszaru zainteresowania widza (ROI – ang. *region of interest*), zaś trzecia interakcji z elementami graficznego interfejsu użytkownika na podstawie analizy danych okulograficznych (zarządzanie kamerą cyfrową w grze wideo i pozyskiwanie kontekstowej, zależnej od zainteresowania, warstwy opisowej elementów gry i interfejsu). Każde z rozwiązań zostało zaimplementowane i przetestowane przez Autora w osobnym i dobrze zaplanowanym cyklu eksperymentów. Dane zostały opracowane statystycznie i przeanalizowane. Chwili zastanowienia wymagają jedynie przeprowadzone i opisane w rozdziale 2.4.3.3 badania ewaluacyjne gry komputerowej, które zakładały testowanie najpierw

sceny bez wspomagania, a potem ze wspomaganiem metodą okulograficzną. Osobiście wydaje mi się, że bardziej wiarygodny wynik można uzyskać losując kolejność testowania scen, gdyż inaczej mamy do czynienia z pewną tendencyjnością doświadczeń. Mimo stwierdzeń Autora, który przyznaje, że tylko dla części koncepcji udało się wykazać użyteczność zaproponowanych rozwiązań, to pozostałe przypadki stanowią bardzo cenne źródło wiedzy o użyteczności mechanizmów postrzegania w interakcji człowieka z komputerem.

Wnioski wynikające z doświadczeń związanych z opartym na uwadze widza, interaktywnym zarządzaniem rozmyciem obrazu, symulującym jego uproszczoną głębię ostrości, zwróciły uwagę Autora na problemy towarzyszące postrzeganiu realizmu sceny przez użytkowników. W pracy została zaproponowana **autorska metoda interaktywnej, opartej na danych okulograficznych, symulacji zjawiska głębi ostrości**. Nowością metody jest dokładniejsze niż w innych przypadkach szacowanie płaszczyzny ostrzenia w obserwowanym środowisku wirtualnym. W pierwszym eksperymencie informacje te pozyskiwane były z wideookulografu jednoocznego i interpretującej je metody I-DT. Oszacowana dzięki nim względna pozycja obiektu przykuwającego uwagę widza stosowana była do określania płaszczyzny ostrzenia w scenie (tzw. płaszczyzna zerowej paralaksy). Rozmycie zaś było realizowane w przestrzeni obrazu, dzięki odpowiedniej implementacji algorytmu na karcie graficznej. Eksperymenty wykazały zauważalną preferencję użytkowników względem interaktywnego zarządzania głębią ostrości sceny trójwymiarowej. Jedyną kwestią, która może zastanawiać jest fakt, że Autor nie użył do szacowania obiektu zainteresowania swojej metody GDOT. W kolejnej odsłonie przeprowadzono eksperyment, w którym weryfikowana metoda interaktywnej symulacji głębi ostrości, szacowała płaszczyznę ostrzenia na podstawie danych z okulografu binokularnego, na podstawie konwergencji oczu. Wykazano, że wprowadzona interaktywna głębia ostrości niweluje zmęczenie i ból głowy wynikający z dobrze znanego konfliktu akomodacji i konwergencji, który obniża komfort stereoskopowej obserwacji wirtualnych środowisk.

W rozdziale 3.2 Autor opisuje eksperymenty mające na celu weryfikację wpływu kontrastu na widzenie głębi i wrażenie realizmu prezentowanej sceny. Badania zostały przeprowadzone w kontekście stereoskopowego widzenia. Profesjonalnie przygotowany i przeprowadzony eksperyment dowiódł wpływu kontrastu i luminancji na postrzeganie głębokości obiektów cyfrowej sceny. Te wskazówki głębi są jednak znacznie słabsze niż wskazówki obuoczne np.: konwergencja. Kolejny eksperyment wykazał również preferencję wyższego kontrastu, jako czynnika wpływającego na realizm oglądanego przez użytkowników obrazu sceny syntetycznej.

4. Na czym polega oryginalny dorobek autora i jakie jest jego znaczenie poznawcze lub przydatność praktyczna dla nauki bądź techniki?

Główne dokonania doktoranta skupiają się wokół obszaru spersonalizowanego zwiększenia ilości informacji przestrzennych w syntezy obrazach cyfrowych. Autor dokonał nie tylko przeglądu i analizy różnych mechanizmów zwiększających realizm obrazów cyfrowych, ale również opracował i zaimplementował szereg metod i technik, które poprawiają fotorealizm postrzeganego obrazu poprzez, sprofilowane pod kątem użytkownika, zwiększenie spektrum bodźców, zwykle deficytowych w syntetycznych obrazach. Dostarczanie bodźców odbywało się interaktywnie (w czasie zbliżonym do rzeczywistego) i zależnie od punktu koncentracji wzroku obserwatora. Tym samym dostarczanie bodźców przestrzennych odbywało się w sposób efektywny i dostosowany do oczekiwań odbiorcy.

Autor podbudował swoje dokonania szeregiem badań o charakterze podstawowym, które szczegółowo zostały zestawione we wnioskach końcowych pracy (strona 138). Wśród nich do najważniejszych należy zaliczyć:

- eksperymentalne określenie relacji pomiędzy kontrastem a postrzeganiem odległości obiektów stosując wyświetlacze o zwiększonej dynamice luminancji;

- określenie związku pomiędzy występowaniem obrazowania zjawiska głębi ostrości, a wrażeniem głębi i zanurzenia w trójwymiarowej scenie;

Badania o charakterze podstawowym były możliwe dzięki opracowaniu autorskiej metody interaktywnej symulacji głębi ostrości, stosującej dane pochodzące z okulografii. W ramach prac Autor stworzył własny, niskobudżetowy system śledzenia ruchu gałki ocznej, którego jakość nie ustępowała profesjonalnym urządzeniom. Opracował również autorską metodę GDOT, której zadaniem było zwiększenie dokładności identyfikacji elementów obrazu, na których spoczywa uwaga obserwatora.

Szczególne znaczenie poznawcze dostarczają wnioski z badań o charakterze podstawowym. Za szczególnie istotne uważam określenie związku pomiędzy zjawiskiem głębi ostrości a niwelacją skutków konfliktu akomodacji i konwergencji, wpływającego istotnie na obniżenie komfortu i zanurzenia w środowiskach wirtualnych. Opracowane przez Autora niskobudżetowe stanowisko okulograficzne, uzbrojone w metodę skutecznego szacowania obiektu zainteresowania widza jest w stanie przyczynić się do dalszego rozwoju badań w tym obszarze i dostarczyć wielu użytecznych rozwiązań dla psychologów, lekarzy, czy osób niepełnosprawnych. Zrealizowane w pracy eksperymenty i wyciągnięte z nich wnioski świadczą dobitnie o wkładzie Autora w poruszane obszary dyscypliny informatyka.

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

Autor istotną część pracy poświęcił przeglądowi zjawisk i mechanizmów towarzyszących postrzeganiu wzrokowemu człowieka. Dokonany przegląd literaturowy jest wszechstronny i dobrze dobrany, chociaż został on wielokrotnie ograniczany, a wnikliwie zostały przedyskutowane tylko najważniejsze z punktu widzenia badań aspekty. Bibliografia składa się ze 155 pozycji, nie licząc 10 publikacji, których Pan Bartosz Bazyluk jest autorem lub współautorem. Dobór i sposób analizy literatury świadczy jednoznacznie o wszechstronnej wiedzy Autora z obszaru interakcji człowiek-komputer, grafiki komputerowej i dziedzin pokrewnych.

Na uwagę zasługuje fakt, że właściwie każde z głównych osiągnięć pracy zostało opublikowane i tym samym uznane przez kolegia redakcyjne uznanych międzynarodowych czasopism i konferencji. Spośród artykułów jedna została opublikowana w czasopiśmie z listy A (*Computer Graphics Forum*) i wiele z pozostałych prac Autora jest indeksowana w bazie *Web of Science*. Jedyna wątpliwość dotyczy braku publikacji autora w okresie trzech ostatnich lat, co trudno jest wyjaśnić postrzegając jego potencjał badawczy, którego dowodem jest zaprezentowana dysertacja.

6. Jakie są wady i słabe strony rozprawy?

Praca nie zawiera istotnych wad, ani znaczących uchybień. Nasuwające się spostrzeżenia mają jedynie charakter dyskusyjny i nie podważają mojej pozytywnej opinii o całości rozprawy.

- Eksperymenty mające na celu weryfikację jakości algorytmu GDOT, przeprowadzane dla syntetycznych środowisk komputerowych, dość marginalnie potraktowały kwestię ograniczeń (założeń) związanych ze złożonością przestrzenną obiektów, prędkością ich poruszania, liczbą obiektów w wyświetlanej scenie. Proszę o doprecyzowanie tych założeń i określenie ich wpływu na skuteczność oraz wydajność metody GDOT.

- Badania w zakresie realizmu postrzegania stereoskopowych obrazów syntetycznych, zrealizowane w ramach rozprawy doktorskiej, przeprowadzane zostały przy użyciu autorskich i komercyjnych wideookulografów rejestrujących ruchy gałek ocznych i syntezujących obraz w odległości kilkudziesięciu centymetrów od twarzy obserwatora. Czy były prowadzone eksperymenty dotyczące użyteczności opracowanych rozwiązań dla konsumenckich wyświetlaczy wirtualnych środowisk typu HMD (ang. *head mounted display*) (Oculus Rift, HTC Vive, VR Gear), w których obraz syntezowany jest kilka centymetrów od

oczu i czy korzystano z dedykowanych im systemów śledzenia wzroku? Jakie są perspektywy realizacji takich badań?

- W rozprawie parokrotnie pojawia się pojęcie wirtualnego środowiska bez zamieszczenia definicji tego terminu. Można odnieść wrażenie, że pojęcie to jest używane zamiennie z graficznym środowiskiem komputerowym, choć w moim odczuciu nie są to pojęcia tożsame. Proszę o skomentowanie tej kwestii.

- W analizie wektorów prędkości metody GDOT pojawia się zastrzeżenie, że ze względu na zaszumienie sygnału, Autor bierze pod uwagę jedynie kąt, a nie długość, pomiędzy chwilowymi wektorami prędkości punktu skupienia i obiektu. Nie zamieszczono stosownej analizy błędów (szumu sygnału), które skłoniły Autora do takiej decyzji. Czy założenie jest poprawne dla sceny, w której wiele obiektów porusza się z różnymi prędkościami?

Praca jest przygotowana starannie pod względem edycyjnym, z zamieszczonym spisem literatury, ale nie zamieszczono w niej spisu rysunków, tabel, ani symboli. Układ pracy jest czytelny. W dysertacji można spotkać bardzo nieliczne błędy literowe i interpunkcyjne (np.: str. 44 „w w punkty”, zamiast „w punkty”, str. 8,13,15, ... brak przecinka przed „na których/w którym”, itp.). Wśród drobnych niedopatrzeń należy wymienić fakt, że w tabeli 2.1 brakuje opisu jednostek, w których podawane są wartości liczbowe, na stronie 88 pojawia się odwołanie do rysunku 5, podczas gdy Autorowi prawdopodobnie chodziło o Rys. 2.5. Dyskusyjny jest brak numeracji wzorów i równań w pracy, które co prawda są dobrze wyjaśnione i osadzone w treści pracy, ale trudno się do nich odwołać w innych miejscach tekstu rozprawy. Wątpliwości budzi również fakt, że niektóre rysunki (np.: 2.18, 2.23, 2.26, 2.27, 2.29, 2.32) mają opisy w języku angielskim podczas gdy cała rozprawa jest w języku polskim. Wydaje się, że dla zachowania spójności przekazu wszystkie elementy rozprawy powinny być przetłumaczone na język polski.

7. Wniosek końcowy

Bardzo dobrze oceniam poziom merytoryczny rozprawy. Autor wykazał się obszerną wiedzą w dziedzinie modelowania percepcji wzorkowej i metod syntezy obrazów cyfrowych. Opracował oryginalne i skuteczne metody szacowania obszarów (obiektów) zainteresowania widza oraz wykonał szereg badań eksperymentalnych o charakterze poznawczym, udowadniając część z postawionych hipotez naukowych i dowodząc tezy rozprawy.

Stwierdzam, że Autor wniósł istotny wkład do dyscypliny naukowej informatyka w zakresie grafiki komputerowej oraz komunikacji człowiek-komputer.

Uważam, że przedłożona do recenzji rozprawa mgr inż. Bartosza Bazyluka z nawiązką **spełnia** wymagania Ustawy o stopniach naukowych i tytule naukowym z 14 marca 2003 roku, Dziennik Ustaw Nr 65, poz. 595 (z późniejszymi zmianami) odnośnie stopnia doktora nauk technicznych. Wnioskuje również o **wyróżnienie rozprawy**.

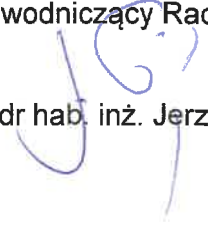
Adam Wojciechowski

UCHWAŁA nr 14/2017/2018
Rady Wydziału Informatyki
Zachodniopomorskiego Uniwersytetu Technologicznego w Szczecinie
z dnia 5 grudnia 2017 r.

**w sprawie nadania stopnia doktora nauk technicznych
w dyscyplinie Informatyka mgr inż. Bartoszowi Bazylukowi**

Zgodnie z Ustawą 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 późniejszymi zmianami) oraz ze zmianami w Ustawie z dnia 27.07.2005 r. *Prawo o szkolnictwie wyższym* Rada Wydziału Informatyki na posiedzeniu w dniu **5.12.2017r** w wyniku głosowania tajnego podejmuje Uchwałę o nadaniu stopnia doktora nauk technicznych w dyscyplinie *Informatyka* mgr inż. Bartoszowi Bazylukowi.

Przewodniczący Rady Wydziału


dr hab. inż. Jerzy Pejaś