

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie  
Wydział Informatyki

METODY MODELOWANIA  
PERCEPCJI GŁĘBI ORAZ WIDZENIA KIERUNKOWEGO  
NA POTRZEBY OBRAZOWANIA KOMPUTEROWEGO

Bartosz Bazyluk

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

Szczecin 2017

# Spis treści

<b>Wprowadzenie</b>	<b>3</b>
Definicja problemu . . . . .	3
Teza i cele pracy . . . . .	4
Metodologia badań . . . . .	5
Organizacja treści rozprawy . . . . .	6
<b>1 Realizm, głębia i uwaga wzrokowa w postrzeganiu obrazu</b>	<b>7</b>
1.1 Postrzeganie obrazu . . . . .	7
1.1.1 Realizm obrazu . . . . .	9
1.1.2 Wskazówki głębi . . . . .	12
1.2 Uwaga wzrokowa . . . . .	16
1.2.1 Pojęcie uwagi wzrokowej . . . . .	16
1.2.2 Fizjologia kierunkowości widzenia . . . . .	17
Podsumowanie rozdziału . . . . .	21
<b>2 Modelowanie kierunkowości widzenia</b>	<b>22</b>
2.1 Okulografia . . . . .	23
2.1.1 Kierunek spojrzenia . . . . .	23
2.1.2 Eye tracking, gaze tracking i okulografia . . . . .	24
2.1.3 Metody okulograficzne . . . . .	24
2.1.4 Mapowanie danych okulograficznych . . . . .	28
2.1.5 Budowa prostego okulografu . . . . .	36
2.2 Problem jakości danych . . . . .	40
2.2.1 Czynniki wpływające na jakość . . . . .	40
2.2.2 Miara jakości systemu okulograficznego . . . . .	43
2.2.3 Pomiar jakości . . . . .	44
2.2.4 Filtracja danych . . . . .	52
2.3 Metoda GDOT . . . . .	56
2.3.1 Algorytm identyfikacji . . . . .	56
2.3.2 Ocena skuteczności identyfikacji . . . . .	64
2.4 Zastosowania metody GDOT . . . . .	73
2.4.1 Analiza uwagi osób oglądających wideo . . . . .	73
2.4.2 Przyspieszenie syntezy obrazu . . . . .	76
2.4.3 Interfejsy użytkownika oparte na danych okulograficznych . . . . .	83
Podsumowanie rozdziału . . . . .	91



<b>3 Modelowanie widzenia głębi</b>	<b>92</b>
3.1 Interaktywna symulacja głębi ostrości . . . . .	92
3.1.1 Określenie płaszczyzny ostrzenia . . . . .	95
3.1.2 Symulacja rozmycia obrazu związanego ze zjawiskiem głębi ostrości . . . . .	96
3.1.3 Badanie skuteczności . . . . .	109
3.1.4 Niezgodność konwergencji i akomodacji . . . . .	114
3.2 Głębia w obrazach o poszerzonym zakresie dynamiki luminancji . . . . .	121
3.2.1 Wpływ kontrastu na postrzeganie głębi . . . . .	123
3.2.2 Kontrast i głębia a poczucie realizmu . . . . .	128
Podsumowanie rozdziału . . . . .	136
 <b>Podsumowanie</b>	 <b>137</b>
Wnioski końcowe . . . . .	138
Odniesienie do tezy . . . . .	139
Dalsze prace . . . . .	139
 <b>Bibliografia</b>	 <b>141</b>
Publikacje własne . . . . .	141
Literatura . . . . .	143

# Wprowadzenie

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 objekty. Naturalnym więc jest, że cechujące każdego widza subiektywne *poczucie realności* otoczenia (ang. *realism perception*), wymaga od obrazu naśladującego rzeczywisty świat, by ten stworzył 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.

## Definicja problemu

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. 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 zróżni-

cowanych wrażeń obuocznych 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 naśladowujący 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 przewyciężyć te ograniczenia oraz wykorzystać dostępne środki tak, aby jak najdoskonalej przekazać niezbędną informację o przestrzenności sceny.

## Teza i cele pracy

Odnosząc się do opisanej wyżej problematyki, na potrzeby rozważań ujętych w niniejszej 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.

Do szczegółowych celów niniejszej rozprawy zalicza się:

- Zbadanie skuteczności metod filtracji danych okulograficznych na potrzeby identyfikacji elementu, na którym spoczywa uwaga obserwatora w środowiskach trójwymiarowych (zob. Rozdział 2.2.4) i zaproponowanie metody dedykowanej dynamicznym wizualizacjom wirtualnych środowisk (zob. Rozdział 2.3).

- Zbadanie możliwych zastosowań wiedzy o punkcie skupienia wzroku w wizualizacji wirtualnego środowiska zarówno w kontekście możliwości przyspieszenia renderowania (zob. Rozdział 2.4.2), jak i elementów interfejsu użytkownika (zob. Rozdział 2.4.3).
- Opracowanie metody interaktywnej, opartej na danych okulograficznych symulacji zjawiska głębi ostrości, zbadanie jej wpływu na odbiór wizualizacji wirtualnego środowiska przez użytkownika (zob. Rozdział 3.1) oraz zbadanie możliwości zredukowania za jej pomocą dyskomfortu konwergencji i akomodacji w wyświetlaczach stereoskopowych (zob. Rozdział 3.1.4).
- Zbadanie wpływu kontrastu w stereoskopowych wyświetlaczach o poszerzonym zakresie dynamiki luminancji na postrzeganie głębi (zob. Rozdział 3.2.1) oraz wrażenie realności oglądanego obrazu (zob. Rozdział 3.2.2).

## Metodologia badań

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 będące-

go 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.

## Organizacja treści rozprawy

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 poczucia 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.

Cała praca stanowi więc 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.

# Rozdział 1

## Realizm, głębia i uwaga wzrokowa w postrzeganiu obrazu

W rozdziale przedstawiono kognitywne oraz fizjologiczne podstawy poczucia realizmu oglądanego obrazu, postrzegania głębi oraz mechanizmu kierunkowości widzenia człowieka. Przybliżone zostają także powiązania pomiędzy poszczególnymi aspektami tych zagadnień, ze szczególnym uwzględnieniem obszaru obrazowania komputerowego.

### 1.1 Postrzeganie obrazu

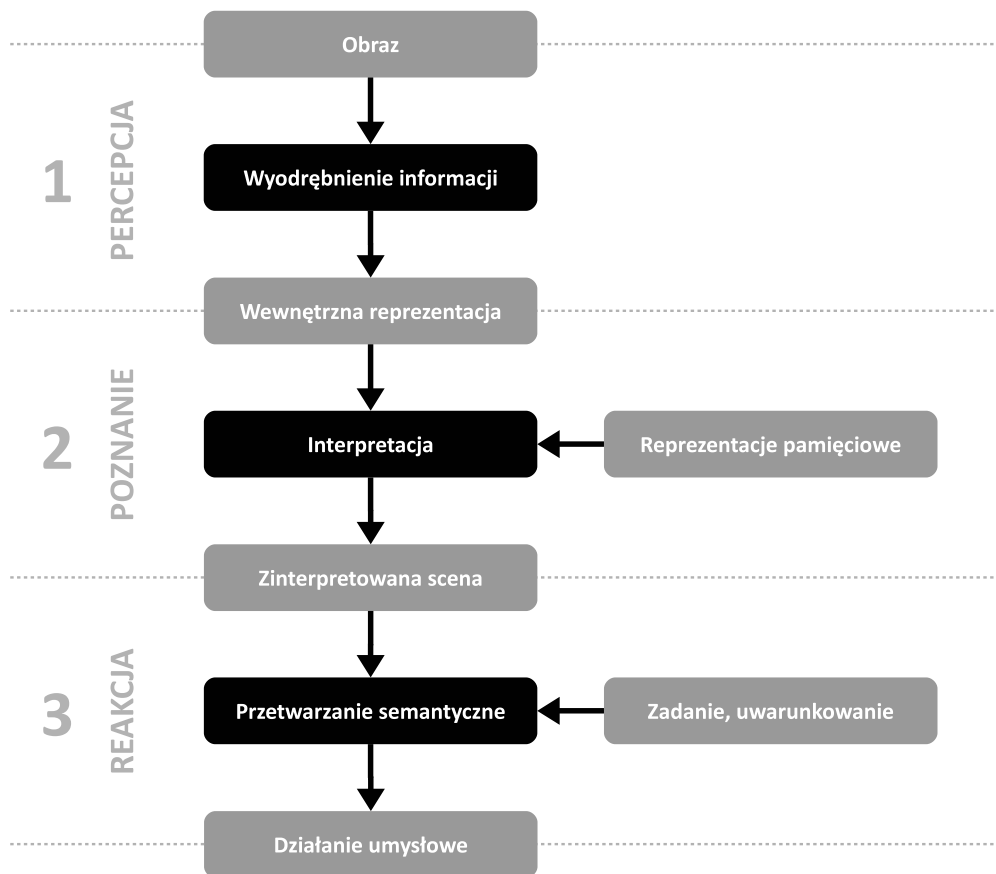
Rozumienie poznawcze obrazu jest procesem kognitywnym, służącym przetwarzaniu zawartej w nim informacji wizualnej. Przyjęło się rozpatrywać ten proces na trzech zaproponowanych przez Marra poziomach abstrakcji: implementacyjnym (*fizycznym*), algorytmicznym (*reprezentacyjnym*) oraz semantycznym (nazywanym również *obliczeniowym*) [Mar82]. Współcześnie wyróżnia się także dwa wyższe poziomy związane z uczeniem i ewolucją [Pog12].

W ujęciu semantycznym, które będzie stosowane w niniejszej rozprawie, postrzeganie obrazu prowadzi do skonstruowania świadomości informacji zawartych w wizualnym bodźcu, by w umyśle obserwatora mogła zajść reakcja na jego oglądanie. W procesie tym można wyróżnić etapy, które prowadzą do uzyskania reprezentacji informacji na kolejnych poziomach konstruowania świadomości (zob. Rys. 1.1) [JB97].

Pierwszy z etapów, *percepcja obrazu*, obejmuje proces transdukcji oraz tworzenia wewnętrznej reprezentacji bodźca, zachodzi zwykle poza świadomością i jest podatny na liczne ograniczenia fizjologii aparatu widzenia. Jest to przetwarzanie niskopoziomowe, operujące w dziedzinie światła i tego, jak w wyniku pobudzenia za jego pomocą fotoreceptorów układ nerwowy jest w stanie utworzyć wstępną reprezentację informacji w postaci trójwymiarowych kształtów i ich hierarchii. Reprezentacja ta stanowi dane wejściowe dla dalszego rozpoznania i klasyfikacji elementów środowiska wizualnego [Wat88, Mar82].

Kolejnym etapem jest *interpretacja* utworzonej przestrzennej struktury informacyjnej, polegająca na wykorzystaniu zgromadzonych wcześniej w pamięci reprezentacji opartych na doświadczeniu. Tak powstała koncepcja sceny jest następnie

Rysunek 1.1: Kognitywny proces postrzegania obrazu w ujęciu semantycznym. Można wyróżnić trzy etapy przetwarzania informacji: niskopoziomową percepcję, zależną od doświadczenia interpretację, a także uzależnione od chwilowego uwarunkowania przetwarzanie semantyczne [JB97].



przetwarzana na poziomie *semantycznym*, w sposób zależny od rozwiązywanego zadania czy też chwilowego uwarunkowania [JB97, dG00].

Mając na uwadze powyższą, semantyczną strukturę procesu postrzegania obrazu, zadaniem syntezy i wyświetlania komputerowej wizualizacji wirtualnej rzeczywistości staje się zapewnienie, by rezultat procesu jej postrzegania odpowiadał rezultatowi, który miałby miejsce, gdyby zamiast na obraz, widz spoglądał na rzeczywiste otoczenie.

### 1.1.1 Realizm obrazu

Jednym z głównych kierunków rozwoju metod obrazowania komputerowego jest dążenie do uzyskania zdolności komunikowania odbiorcy takich obrazów, które byłyby w stanie naśladować rzeczywistość. Obejmuje to zarówno metody samej syntezy, jak i prezentacji obrazu. Niezależnie jednak od ich doskonałości i zaawansowania, wizualizowany świat jest jedynie reprezentacją obiektów, które są jego częścią. Trudną do zdefiniowania miarę skuteczności naśladowania faktycznych obiektów przez tę wizualizację, zwykło się nazywać *realizmem* danego obrazu (zob. Rys. 1.2).

Z realizmem obrazu nieodłącznie powiązane jest także pojęcie jego jakości. Za Janssenem i Blommaert, na potrzeby dalszych rozważań przyjęte zostaną dwa sposoby definicji tego pojęcia [JB97]. Pierwszy z nich odnosi się do niedoskonałości i ograniczeń samego sygnału, którym jest obraz. Drugi ze sposobów definicji jakości dotyczy zaś informacji, która jest komunikowana przez dany obraz. W tym ujęciu obrazem *dobrej jakości* nazywa się ten, który może zostać skutecznie zinterpretowany przez odbiorcę, a więc pozwala mu odtworzyć niesioną przez ten obraz informację. W kontekście wirtualnych środowisk będzie to więc uzyskanie przez użytkownika pełni świadomości sytuacyjnej na temat świata, który jest wizualizowany. W niniejszej pracy, wiodące pojęcie jakości obrazu będzie odnosiło się do ostatniego z przytoczonych podejść.

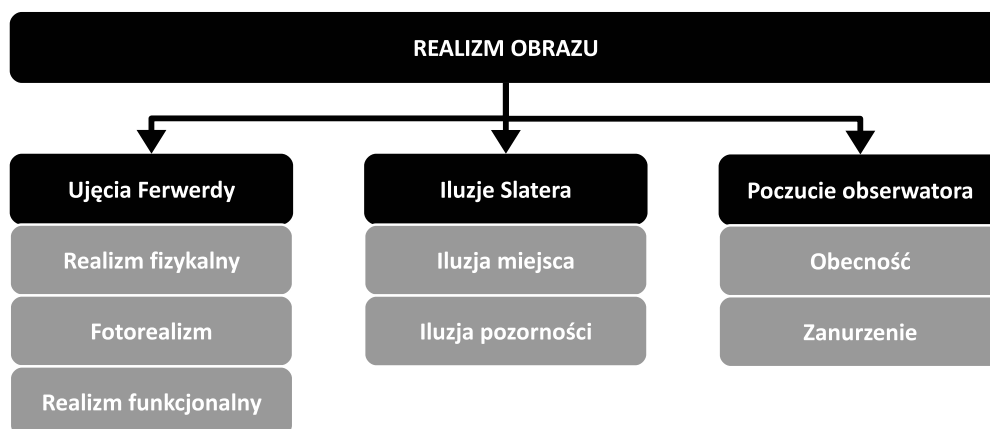
#### 1.1.1.1 Ujęcia realizmu obrazu komputerowego

James Ferwerda w swojej często cytowanej pracy na temat realizmu obrazów komputerowych [Fer03], wyróżnia trzy ujęcia tego pojęcia, różniące się płaszczyznami na których odbywa się naśladowanie rzeczywistości: *realizm fizyczny* (ang. *physical realism*), *fotorealizm* (ang. *photo-realism*) i *realizm funkcjonalny* (ang. *functional realism*).

Mówiąc o *realizmie fizycznym*, za obraz *realistyczny* uznaje się taki, który dokonuje stymulacji aparatu widzenia obserwatora dostarczając dokładnie takich samych wartości energii, jakich dostarczałyby same znajdujące się na nim obiekty. Ujęcie to wymaga więc zarówno dokładnych obliczeń prowadzących do uzyskania koniecznych wartości, jak i sposobu wyświetlania będącego w stanie dostarczyć obserwatorowi niezbędnej stymulacji. Należy jednak pamiętać, że nawet najdoskonalszej jakości, komputerowo syntezowana statyczna wizualizacja sceny, nie jest w stanie w uniwersalny sposób spełnić kryterium realizmu fizycznego, gdyż wyjście systemu komputerowego posiada ograniczoną zdolność do reprodukcji cech obrazu (ograniczony kontrast, ograniczone zdolności prezentacji obrazu panoramicznego czy obuocznego). Wyzwaniem jest więc dostarczenie ludzkiemu systemowi widzenia takich bodźców,



Rysunek 1.2: Różne podejścia do definicji realizmu obrazu.



które wywołają wrażenie głębi *zbliżone* do obserwacji sceny w świecie rzeczywistym. Pewnym ułatwieniem może być fakt, że idea *realizmu fizykalnego* nie bierze pod uwagę ograniczeń aparatu widzenia, dostarczając jedynie źródłowego bodźca, nawet jeśli ten z powodu wspomnianych ograniczeń nie będzie w pełni odbierany. To tworzy możliwość ominięcia niektórych z limitów współczesnych systemów wyświetlania obrazu i toruje drogę do kolejnego z ujęć realizmu grafiki komputerowej.

Uwzględniając cechy aparatu widzenia człowieka i stawiając sobie za cel jedynie wywołanie takiej samej *reakcji wzrokowej* jak podczas oglądania rzeczywistej sceny, mówimy o pojęciu *fotorealizmu*. To ujęcie realizmu obrazu komputerowego ogranicza konieczność dokładnej reprodukcji całości bodźca, a więc na przykład fizycznych poziomów energii docierającej do oka, skupiając się jedynie na naśladowaniu samego *wrażenia*. Pozwala więc na mówienie o obrazach realistycznych nawet przy ograniczeniach technicznych, jakimi obarczone są współczesne graficzne systemy komputerowe: zarówno w kontekście zdolności do reprodukowania samego obrazu, jak i wydajności obliczeniowej. Wieloaspektowość tego zagadnienia obejmuje szerokie spektrum elementów związanych z ludzką percepcją wzrokową: postrzegania koloru, kontrastu, układu przestrzennego czy głębi.

Natomiast pojęcie *realizmu funkcjonalnego* dotyczy skupienia się na odwzorowaniu istotnych w danej chwili zależności i zachowań obrazowanego wycinka świata. Obrazy realistyczne funkcjonalnie naśladowują więc niesienie tej samej *informacji*, co obserwacja rzeczywistego pierwowzoru. Tu wyróżnić należy, szczególnie interesującą z uwagi na przedmiot rozprawy, zależność postrzeganego obrazu od naturalnych zachowań obserwatora. Przemieszczenie punktu skupienia wzroku na inny obiekt wywołuje zmianę bodźca pobudzającego siatkówkę w wyniku akomodacyjnego dostosowania się kształtu soczewki oka do odległości tego obiektu. Podobnie jasność oglądanego, rzeczywistego obiektu może powodować reakcję w postaci adaptacji sensorycznej do warunków oświetleniowych. Te, charakterystyczne dla obserwa-

cji rzeczywistego świata zjawiska nie zachodzą, gdy wizualizujemy scenę na klasycznym wyświetlaczu o ograniczonej zdolności do reprodukcji dynamiki luminancji. By dążyć do osiągnięcia realizmu funkcjonalnego, należy więc te i podobne zjawiska imitować już na etapie syntezy obrazu.

W niniejszej pracy poruszana będzie kwestia realizmu rozumianego w każdym z trzech opisanych wyżej ujęć. Realizm fizyczny zostanie ujęty podczas badania na wyświetlaczach o poszerzonym zakresie dynamiki wpływu kontrastu luminancji na ocenę obrazu (zob. Rozdz. 3.2.2), fotorealizm będzie dominującym z aspektów podczas próby naśladowania wrażenia głębi poprzez różne techniki syntezy obrazu (zob. Rozdz. 3.2.1), zaś w kontekście realizmu funkcjonalnego zostanie rozpatrzona interaktywna symulacja głębi ostrości i jej wpływ na złagodzenie dyskomfortu podczas prezentacji obrazów stereoskopowych (zob. Rozdz. 3.1.3 i Rozdz. 3.1.4).

### 1.1.1.2 Zanurzenie, obecność, iluzja miejsca i pozorności

Główną ideą wirtualnych środowisk i działania systemów graficznych mających je wizualizować, jest nie tyle przedstawienie użytkownikowi samego obrazu, co za jego pomocą wywołanie w użytkowniku poczucia uczestniczenia w funkcjonowaniu obrazowanego świata [SW97].

Slater w swoim artykule [Sla09] zwraca uwagę, że graficzne systemy komputerowe tworzące wizualizację wirtualnej rzeczywistości nie służą wyłącznie naśladowaniu realnych obrazów. Ich zadaniem jest wytworzenie poczucia realizmu tej sceny, której częścią staje się użytkownik, jednak nie musi to być związane z dokładnym odzwierciedleniem znanego, otaczającego świata. W pracy tej wprowadzone zostaje rozróżnienie dwu pojęć dotyczących odczuć użytkownika wirtualnej rzeczywistości: iluzji miejsca *PI* (ang. *place illusion*) oraz iluzji pozorności *Psi* (ang. *plausability illusion*).

O iluzji miejsca mówimy, gdy użytkownik odczuwa dostarczane przez wirtualną rzeczywistość bodźce jako prawdziwe. Ma więc poczucie obecności w miejscu, które jest przez nią symulowane. Można to nazwać realizmem wysokiego poziomu, głębokim i porównywalnym z fizycznym znajdowaniem się w wirtualnym świecie.

Natomiast iluzja pozorności dotyczy przekonującego dostarczenia symulacji zachowań świata, niekoniecznie powiązanego z przedstawieniem obrazów identycznych do tych, które użytkownik oglądałby w świecie realnym. Jest to więc płytsza forma realizmu, nawiązująca przede wszystkim do realizmu funkcjonalnego Ferwerdy.

Powyższe prowadzi do dwu kolejnych ważnych konceptów: poczucia *obecności* użytkownika w wirtualnej rzeczywistości oraz jego *zanurzenia* w tejże. Pierwsze z nich definiowane jest jako stopień, w jakim system postrzegania i kognitywnego przetwarzania informacji daje się ponieść złudzeniu przebywania w innej lokacji niż ma to fizycznie miejsce [BC04]. Ten trudny do osiągnięcia stan [PCS<sup>+</sup>00] jest więc zależny od subiektywnej, psychologicznej reakcji danej osoby i jest charakterystyczny dla danej jednostki [BM07]. Z drugiej strony, zanurzenie jest definiowane jako rzecz obiektywna, możliwa do pomierzenia i zależna od zdolności danego systemu wirtualnej rzeczywistości do pobudzenia reakcji psychofizycznej u użytkowników, porównywalnej z tą, która zachodzi przy eksploracji realnego świata. Jest ona więc zależna od kompletności spektrum informacji, które dany system przekazuje odbiorcy [Sla03].

### 1.1.1.3 Komunikowanie informacji przestrzennej

By wywołać u odbiorcy wizualizacji iluzję pozorności, poczucie obecności i osiągnąć wysokie zanurzenie, konieczne jest zadbanie o zdolność wytworzenia wrażenia uczestnictwa w funkcjonującym, przestrzennym środowisku. Takim, które dostarcza nie tylko odczuć estetycznych, lecz pozwala na świadome określenie związków pomiędzy tworzącymi je obiektami, a samym użytkownikiem, w tym związków opartych na odległości. Konieczne dla osiągnięcia akceptowalnego realizmu jest więc komunikowanie odbiorcy informacji o głębi w sposób możliwie wiarygodny.

### 1.1.2 Wskazówki głębi

Człowiek wytwarza świadomość przestrzenności otoczenia na podstawie różnorodnych przesłanek wizualnych, postrzeganych w oglądanym obrazie. Przesłanki te zbiorczo nazywa się wskazówkami głębi [ZJM09]. Oddziałują one zarówno na etapie percepcji obrazu, jak i jego interpretacji.

Zasadniczy podział wskazówek głębi polega na ich rozróżnieniu z uwagi na to, czy do niesienia informacji wystarczą bodźce odbierane przez pojedyncze oko (*wskazówki monokularne*), czy też wnioskowanie o głębi następuje w wyniku porównania informacji otrzymanych od obojga oczu (*wskazówki binokularne*) (zob. [ZJM09], s. 84–87).

#### 1.1.2.1 Wskazówki binokularne

Rozbieżność dwuoczną, wywołaną obuocznym oglądaniem obrazów rozsuniętych względem siebie o pewną odległość, jest podstawową wskazówką o odległości oglądanego obiektu od obserwatora. Obie gałki oczne są niezależnie orientowane poprzez działanie mięśni okomotorycznych w taki sposób, by ich osie widzenia przecięły się na powierzchni tego elementu świata, na którym w danej chwili skupiony jest wzrok. Wówczas umysł dokonuje fuzji obu, często różniących się w pewnym stopniu, obrazów [HR95]. Dostarczając więc osobne, różniące się zastosowaną projekcją, obrazy dla lewego i prawego oka podczas wizualizowania wirtualnego środowiska, można wymuszać na obserwatorze reakcję zbliżoną do tej, jaką wzbudziłoby oglądanie rzeczywistej, przestrzennej sceny.

### Wyświetlanie obrazu stereoskopowego

Idea dostarczenia obojgu oczom różnych obrazów w celu wywołania wrażenia głębi nie jest nowa. Już w pierwszej połowie XIX wieku proponowano użycie rozwiązań technicznych, pozwalających przekierować tor optyczny oczu tak, by trafił on na specjalnie przygotowane wersje obrazu, zawierające subtelne różnice imitujące rzeczywiste oglądanie trójwymiarowej przestrzeni z dwu rozsuniętych punktów. Oparty o lustro projekt stereoskopu Wheatstone'a [Whe38] do dnia dzisiejszego jest rozwiązaniem używanym w wielu badaniach nad obrazowaniem stereoskopowym. Dzieje się tak, ponieważ nie powoduje on zniekształceń obrazu i zakłóceń w ich odbiorze, charakterystycznych dla innych technik [Pet14].

Popularnym, prostym i pasywnym sposobem prezentacji obrazu stereoskopowego przeznaczonym także dla druku, jest technika anaglifów. Została ona zaproponowana

również w XIX wieku, przez niemieckiego fizyka Rollmanna [Rol53]. Polega ona na wykorzystaniu dwu różnych filtrów barwnych dla obojga oczu oraz specjalnie spreparowanego stereogramu, składającego się z nałożonej na siebie pary obrazów dla oka lewego i prawego, opartych na innych kolorach. Ta technika ma jednak istotne wady związane ze zmęczeniem wzroku podczas korzystania z filtrów barwnych oraz ograniczoną zdolnością do reprodukcji oryginalnej kolorystyki obrazu.

Możliwe jest też wykorzystanie zjawiska polaryzacji i okularów wyposażonych w dwa filtry polaryzacyjne, których płaszczyzna polaryzacji jest ustawiona względem siebie prostopadle. Poprzez odpowiednie spolaryzowanie światła podczas wyświetlania obrazów przeznaczonych dla oka lewego i prawego, możliwe jest zapewnienie, że tylko to oko, które jest adresatem danego obrazu, jest w stanie go obejrzeć. Metoda ta jest jednak podatna na zjawisko przesłuchu (ang. *crosstalk*), polegające na mieszaniu klatek oka lewego i prawego [Zon07]. Jej stosowanie przy prezentacji obrazu komputerowego wymaga specjalnego sposobu wyświetlania, pozwalającego na różną polaryzację światła dla poszczególnych klatek lub pikseli.

Współcześnie najczęściej stosowaną techniką dla konsumenckich wyświetlaczy stereoskopowych jest natomiast podejście oparte o aktywne okulary migawkowe. Szklane okulary używanych przez użytkownika są wyposażone w naprzemiennie przymykane z dużą częstotliwością migawki, najczęściej zbudowane z ekranów ciekłokrystalicznych. Przy synchronizacji częstotliwości naprzemiennego przymykania migawek dla obojga oczu z wyświetlaniem kolejnych klatek animacji na ekranie, możliwe jest adresowanie poszczególnych z nich do odpowiednich oczu. Rozwiązanie to jest jednak związane z wieloma problemami, między innymi związanymi ze zmęczeniem oczu i występowaniem przesłuchu pomiędzy klatkami [Zon07].

Warto zauważyć, że zakładając użycie gogli wirtualnej rzeczywistości, wyświetlanie osobnych klatek przeznaczonych dla poszczególnych oczu staje się często znacznie łatwiejsze. Nie wymaga ono stosowania żadnej z powyższych technik, jeśli fizyczne rozmiary urządzenia pozwalają na użycie dwóch oddzielnych ekranów lub części ekranu z przeznaczeniem do wyświetlania obrazu dla poszczególnych oczu. Wówczas możliwe jest niezależne wyświetlanie klatek dla oka lewego i prawego.

### 1.1.2.2 Wskazówki monokularne

Pomimo tego, że podstawowym przeznaczeniem widzenia obuocznego jest wnioskowanie o głębi, można wyróżnić także liczne wskazówki o odległości elementów obrazu, które opierają się na widzeniu jednoocznym.

## Perspektywa

Ludzki aparat widzenia bazuje na obrazie uzyskanym w wyniku perspektywicznego rzutowania światła na siatkówkę. Prowadzi to do powstania zjawiska, w którym obiekty bardziej odległe od obserwatora, uzyskują mniejszy obraz aniżeli te bliższe. Doświadczenie obserwatora powoduje więc, że naturalnym jest wnioskowanie o odległości ze względnych rozmiarów obiektów lub zbiegania się tworzonych przez nie linii (ang. *depth from shape*) [ZJM09].

## Akomodacja

Cechą każdego fizycznego układu optycznego złożonego z przynajmniej jednej soczewki, jest towarzyszące mu zjawisko głębi ostrości. Dla danego układu istnieje tylko jedna odległość od soczewki, której obrazowanie w wyniku zogniskowania projekcji na powierzchnię siatkówki pozwala na uzyskanie rezultatu pozbawionego rozmycia. Wszystkie obiekty znajdujące się bliżej tej progowej odległości (tzw. *odległości ostrzenia*) i dalej od niej, będą w efekcie nieostre. Jednak z uwagi na ograniczoną rozdzielczość widzenia, zapisywania lub przetwarzania obrazu, pewien zakres rozmycia jest poniżej progu zauważalności. Tak nieznacznie rozmyty obraz jest odczuwany jako ostry. Zakres odległości powodujących rozmycie poniżej wspomnianego progu, nazywany jest zakresem głębi ostrości [MMN99, TC75].

Ludzki aparat widzenia, podobnie do obiektywów aparatów fotograficznych, ma możliwość zmiany odległości ostrzenia zależnie od potrzeb. O ile dla większości obiektywów odbywa się to poprzez zmianę odległości pomiędzy soczewkami, w przypadku ludzkiego aparatu widzenia zmienia się geometria samej soczewki znajdującej się w oku, prowadząc do zmiany jej ogniskowej. Dzieje się to w wyniku zmiany napięcia mięśnia rzęskowego, będącego częścią tęczówki. Wpływ na zakres głębi ostrości ma również rozmiar źrenicy [CW77].

Związane z akomodacją zjawisko głębi ostrości i wynikające z niej selektywne rozmycie obrazu, występują przy oglądaniu obrazów przestrzennych, gdzie tworzące je obiekty znajdują się na różnych odległościach. Rozmycie to staje się więc wskazówką niosącą informację o zróżnicowaniu dystansów poszczególnych elementów świata [ZJM09]. Z tego powodu rozmycie to jest wykorzystywane w fotografii i kinematografii dla podkreślenia głębi reprodukowanej sceny, a także prowadzenia wzroku oglądającego, sugerując miejsca na których powinien on skupić wzrok, by rozmycie to wydało się naturalnym [Mon00].

## Kontrast

Ciągłe doświadczanie w rzeczywistym świecie niepełnej przejrzystości powietrza, mgły oraz smogu jest przyczyną wytworzenia przekonania o większej odległości od obserwatora tych obiektów, które jawią się jako mające mniejszy kontrast. Zjawisko to, niezwykle często wykorzystywane w malarstwie dla ukazania dystansu na pejzażach, nazywane jest *perspektywą powietrzną* (ang. *aerial perspective*) [OBO94].

Do wskazówek bazujących na kontraście zalicza się także wnioskowanie o względnych odległościach i kształcie przedmiotów na podstawie układu światłocieni (ang. *depth from shading*) [Hee06]. Również obiekty, które cechują się większą ilością emitowanego lub odbijanego światła, są często interpretowane jako bliższe [ZJM09, IKA07].

## Przesłonięcie i ruch obiektów

Obiekty bliższe przesłaniają te, które znajdują się dalej od obserwatora. Zjawisko okluzji, rzeczywistej bądź pozornej (imitowanej jako narzędzie sugerowania kolejności elementów [EML13]) jest więc czynnikiem który wpływa na ocenę względnych odległości pomiędzy obiektami.

Znaczenie okluzji jest także istotne, gdy obserwowana scena zawiera ruchome elementy. Tymczasowe, dynamiczne przesłanianie się ich, pozwala wnioskować o ich względnym położeniu. Zachowanie to w połączeniu z estymacją prędkości poruszania się obiektów pozwala dołączyć do zestawu wskazówek o głębi dodatkowy element: zjawisko *paralaksy* (ang. *parallax, depth from motion*) [ZJM09].

### 1.1.2.3 Istotność wskazówek

Mnogość dostępnych w postrzeganiu wizualnym nośników informacji o głębi sugeruje zastanowienie się nad istotnością i siłą każdego z nich. Z uwagi na techniczne trudności związane z użyciem niektórych z nich podczas syntezy i prezentacji komputerowego obrazu, wykazanie związków pomiędzy poszczególnymi wskazówkami głębi, mogłoby utorować drogę do prób kompensacji braków z pomocą innych spośród nich.

Zespół O'Shea podjął się oszacowania wagi, którą obserwatorzy przypisują informacji o głębi płynącej z kontrastu [OBO94]. Wykazał on, że przy monoskopowych bodźcach o zróżnicowanej luminancji, kontrast jest wystarczająco silną wskazówką by konkurować z wnioskowaniem o odległości z rozmiaru obiektów. Jego siła jednak ulega znaczącemu osłabieniu w sytuacji, gdy dostarczana jest dodatkowo wskazówka stereoskopowa. Ta zależność zostanie zbadana i zamodelowana w Rozdziale 3.2.1.

Z punktu widzenia powiązań pomiędzy wskazówkami, istotny jest też związek pomiędzy wskazówką binokularną, opartą o konwergencję oczu, a głębią ostrości. Występowanie tej pierwszej przy jednoczesnym braku drugiej, uważane jest za podstawową przyczynę dyskomfortu odczuwanego przez użytkowników współczesnych jednopłaszczyznowych wyświetlaczy stereoskopowych. Konflikt konwergencji-akomodacji stanowi poważny problem współczesnej techniki wizualizacji trójwymiarowych środowisk [LFH09, IKT09, HR02]. Próba jego złagodzenia zostanie opisana w Rozdziale 3.1.4.

Easa wraz z zespołem podjęli próbę uszeregowania sposobów prezentacji wielowarstwowych danych przestrzennych pod względem ich skuteczności [EML13]. Zbadane metody uwzględniały kilka monoskopowych wskazówek o głębi, którymi intuicyjnie posługują się obserwatorzy. Przeprowadzony eksperyment pozwolił na wskazanie zróżnicowanego, względnego rozmiaru oraz kontrastu jako tych wskazówek, które w warunkach wyizolowania pozwoliły na największą skuteczność identyfikacji właściwej kolejności wyświetlanych warstw.

Z kolei zespół Helda w swojej pracy uzyskał wyniki pozwalające na stwierdzenie, że istotność poszczególnych wskazówek zależy od obszaru pola widzenia, w którym znajdują się postrzegane obiekty [HCB12]. Według ich ustaleń, w bezpośrednim sąsiedztwie punktu skupienia wzroku, największy udział w sugerowaniu odległości ma wskazówka oparta na konwergencji. Dalej od tego obszaru rośnie znaczenie rozmycia wynikającego z głębi ostrości. W zakresie widzenia bocznego i peryferyjnego, najistotniejsza staje się zaś wskazówka oparta na rozmiarach obiektów.

Wrażenie głębi pochodzące z oglądania wizualizacji wirtualnej sceny powinno odzwierciedlać rzeczywiste odczucia, które płynęłyby z zanurzenia w tej scenie. Interante i Anderson w swoich badaniach opisanych na łamach [IRA06], opartych o gogle wirtualnej rzeczywistości i komputerowy model rzeczywistego otoczenia użytkownika, dowiedli że dostarczając użytkownikowi jedynie część możliwych wskazówek uda-

je się wytworzyć świadomość dystansu otaczającego świata. W ich eksperymencie badani mieli za zadanie wchodzić w interakcję z elementami rzeczywistego pomieszczenia, w którym się znajdowali, oglądając na żywo wyłącznie jego wirtualny odpowiednik. Odwzorowanie rzeczywistości odpowiadało faktycznym wymiarom otoczenia, jednak dostarczany, stereoskopowy obraz, pozbawiony był chociażby rozmycia wynikającego z głębi ostrości oraz wskazówek opartych na kontraście. Postrzeganie głębi przez uczestników było jednak wystarczająco silne, by byli oni w stanie swobodnie poruszać się po pomieszczeniu ze świadomością realnych odległości.

## 1.2 Uwaga wzrokowa

Z widzeniem nieodzownie łączy się pojęcie uwagi wzrokowej (ang. *visual attention*). Aby spróbować zrozumieć jej znaczenie, należy jednak zacząć od rozpatrzenia pojęcia uwagi w szerszym ujęciu, odnoszącym się do ogółu procesów percepcyjnych i kognitywnych.

### 1.2.1 Pojęcie uwagi wzrokowej

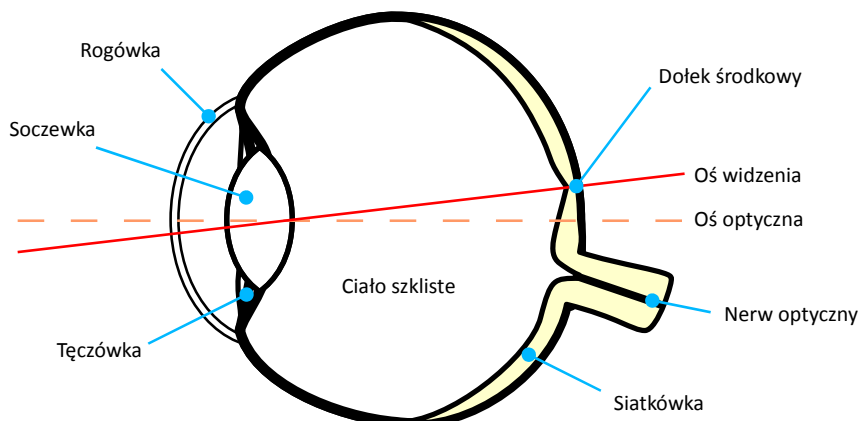
W literaturze można spotkać się ze zróżnicowanym podejściem do tłumaczenia przez badaczy kwestii uwagi. Przykładowo, Broadbent określił jej mianem mechanizm wybiórczego kierowania bodźców do kanałów sensorycznych o przepustowości ograniczonej pewnym nieprzekraczalnym limitem [Bro58].

Z kolei Treisman w swojej teorii integracji cech (ang. *Feature Integration Theory, FIT*), wskazuje na uwagę jako drugi etap procesu postrzegania, poprzedzony odbywającą się w sposób równoległy rejestracją skończonego zbioru cech wielu obiektów. Uwaga jest wówczas wybiórczym wzmacnianiem bądź osłabianiem tych cech lub odnajdywaniem relacji pomiędzy nimi w uzależnieniu od chwilowego ukierunkowania danej osoby. Zaś ostatecznym celem jest wyodrębnienie konkretnego obiektu zainteresowania z całej ich gamy będącej w zasięgu zmysłowym [TG80].

Natomiast Wolfe w serii swoich teorii ukierunkowanego poszukiwania (ang. *Guided Search, GS*), nawiązujących do *FIT*, szczególnie zaakcentował mechanizmy poprzedzające sam proces uwagi. Skupiając się na zadaniu wizualnego odszukania elementu sceny i towarzyszących temu zachowań, wskazał jako kluczowy czynnik mający wpływ na efektywność tego poszukiwania, świadome ukierunkowanie uwagi opierające się zarówno o procesy *top-down* (odgórne, rozpoczynające się od ukierunkowania), jak i *bottom-up* (oddolne, inicjowane samymi bodźcami) [Wol07]. Co istotne, pierwsza wersja *GS* ograniczała się jednak wyłącznie do uwzględniania kierunkowania związanego z procesami kognitywnymi, ignorując specyfikę samych ruchów oczu i budowę siatkówki, które jak się okaże w Rozdziale 2, są niezwykle ważne dla omawianych rozważań (por. [WCF89] oraz [Duc07, s. 229]).

Uwagę można więc próbować interpretować jako zdolność do selektywnego skupiania się na elementach otaczającego świata, pozwalającą odrzucić te z nich, które wydają się być nieistotne przy aktualnym uwarunkowaniu obserwatora. W przeniesieniu na widzenie, oznacza to taką ocenę składników wrażenia wzrokowego, że są one przetwarzane przez umysł w sposób ważony. Należy tutaj podkreślić, że czym innym jest wspomniane ważenie w sensie procesów równoległych poprzedzających

Rysunek 1.3: Uproszczony schemat budowy ludzkiego oka z wyróżnionymi najistotniejszymi jego elementami.



uwagę wizualną w rozumieniu *FIT* Treisman czy *GS* Wolfe'a, czym innym zaś ważenie związane z fizjologiczną budową oka. Mamy więc do czynienia z dychotomią selektywności uwagi wizualnej, o której można mówić zarówno w kontekście kognitywnym, jak i fizjologicznym (zob. Rozdz. 2.1.1).

W znaczeniu czysto fizjologicznym selektywność wynika z niejednorodnej czułości aparatu wzrokowego na detale obrazu z racji budowy siatkówki. Można więc powiedzieć, że następuje w ten sposób naturalne ważenie istotności elementów obrazu, a waga maleje wraz z oddalaniem się od centralnego pola widzenia. Kontrola selektywności w ujęciu fizjologicznym polegałaby zatem na takim ruchu gałek ocznych, by obiekt uwagi znalazł się w obszarze pozwalającym na postrzeganie największej ilości jego szczegółów.

## 1.2.2 Fizjologia kierunkowości widzenia

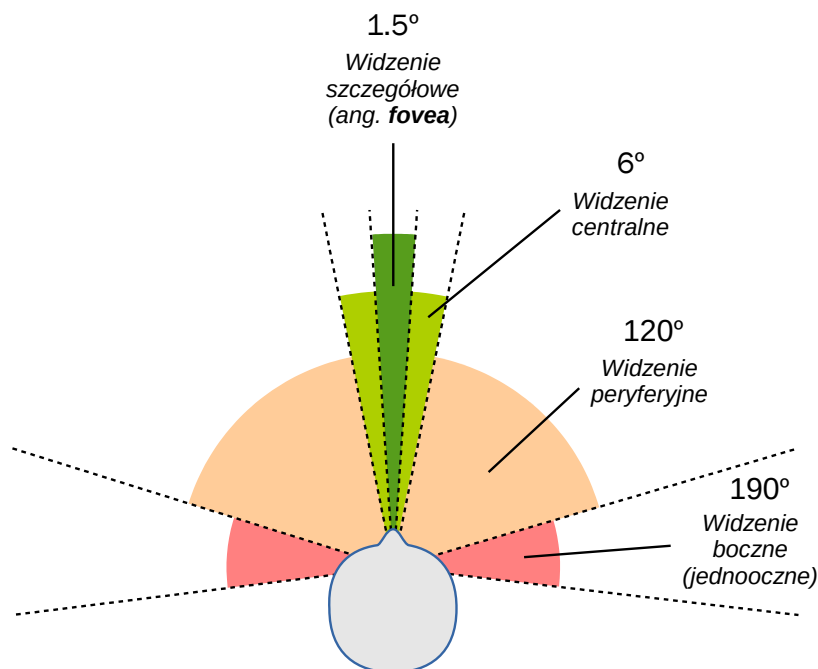
Interesującym zagadnieniem powiązaniem z uwagą wizualną jest kwestia budowy świadomości sytuacyjnej pełni sceny, pomimo ograniczeń chwilowego pola widzenia. U zdrowego człowieka pole widzenia pary nieruchomych oczu przekracza w płaszczyźnie poziomej 180 stopni [Duc07], jednak wyłącznie mały jego wycinek pozwala na widzenie szczegółowe.

### 1.2.2.1 Budowa siatkówki

Elementem realizującym transdukcję, a więc przetwarzającym światło na impulsy nerwowe prowadzące do wytworzenia świadomości obrazu, jest mozaika czopków i pręcików: komórek światłoczułych, czyli fotoreceptorów. Wchodzą one w skład siatkówki wyściełającej dno oka. Ich przestrzennie zróżnicowane zagęszczenie powoduje, że jedynie część obrazu może być postrzegana w sposób wyraźny i szczegółowy [Ost35]. Współcześnie wyróżnia się także dodatkowy rodzaj komórek światłoczułych, komórki zwojowe (ang. *photosensitive ganglion cells*), jednak nie biorą one udziału w tworzeniu podstawowej świadomości oglądanego obrazu, a służą głównie regulacji cyklu dobowego [DY10].



Rysunek 1.4: Kąty widzenia człowieka z rozróżnieniem na widzenie centralne, peryferyjne i boczne. Osoba na rysunku ma wzrok skierowany na wprost, ku górze diagramu.



Środkową część siatkówki z racji intensywnie żółtego pigmentu określa się mianem plamki żółtej. W jej obrębie znajduje się tzw. dołek środkowy oka (łac. *fovea centralis*), który jest obszarem o największym zagęszczeniu czopków odpowiadających za kolorowe widzenie w warunkach fotopowych (dobrego oświetlenia). Oś widzenia, która związana jest z obszarem najwyraźniejszego postrzegania, jest wyznaczana przez odcinek łączący dołek środkowy oka ze środkiem soczewki. Jego przedłużenie wskazuje kierunek w jakim zorientowane jest szczegółowe widzenie i nie należy utożsamiać go z osią optyczną soczewki (ang. *optical axis*), od której jest ona znacznie odchylona (por. Rys. 1.3). Biorąc pod uwagę rozkład fotoreceptorów w plamce żółtej przyjmuje się, że człowiek jest w stanie postrzegać detale w obszarze odchylonym do około  $1^\circ$  od osi widzenia, przy czym zdolność ta jest różna i niejednorodna dla poszczególnych kolorów światła na które czule są czopki [CAS<sup>+</sup>91]. Ta część pola widzenia nazywana jest widzeniem centralnym (ang. *foveal vision*). Poza obszarem centralnym mamy do czynienia z widzeniem peryferyjnym o znacznie obniżonej rozdzielczości i czułości. W jego obrębie można jednak wyróżnić umiarkowane szczegółowe widzenie *parafoveal*, związane z projekcją obrazu na zewnętrzny obszar plamki żółtej. Dotyczy ono obszaru do  $2.5^\circ$  odchylenia od osi widzenia.

Naturalnym jest więc dążenie do tego, by oglądany obiekt znajdował się w obszarze widzenia centralnego, a więc by jego obraz był rzutowany na najbliższą okolicę dołka środkowego. W tym celu wykorzystywane są mięśnie gałki ocznej (tzw. *mięśnie okomotoryczne*), które zmieniają jej orientację. Agregacja wiedzy o całym, złożonym obrazie, który nie może zostać pokryty widzeniem centralnym jest więc sekwencją zbierania części informacji wizualnej pochodzących z jego fragmentów.

Proces oglądania polega na ciągłym przenoszeniu przez obserwatora swojej zdolności wyraźnego widzenia pomiędzy znajdującymi się na obrazie regionami zainteresowania [NS71, Duc07].

### 1.2.2.2 Fiksacje

Moment, w którym stacjonarny względem obserwatora obiekt zainteresowania jest rzutowany na część siatkówki umożliwiającą wyraźne widzenie w taki sposób, by jego projekcja pozostała jak najbardziej nieruchoma, nazywany jest fiksacją. Aby obraz obiektu mógł być widziany wyraźnie, szybkość jego przemieszczania się musi być mniejsza niż  $5^\circ/\text{s}$  [LZ99, s. 189–190]. To właśnie wówczas jest świadomie postrzegana pochodząca z niego, szczegółowa informacja wizualna. Fiksacje zazwyczaj trwają 150–600 milisekund i stanowią 90% czasu oglądania statycznej sceny [Duc07, s. 46–47].

Podczas trwania fiksacji gałka oczna nie pozostaje jednak nieruchoma. Możliwe jest zauważenie nieświadomych ruchów pod postacią niewielkiego drżenia (ang. *tremor*), dryfu (ang. *drift*) oraz tzw. ruchów mikrosakadycznych (ang. *microsaccades*). Natura dwóch pierwszych z nich nie jest oczywista i wiąże się ze swojego rodzaju szumem naturalnego aparatu sterowania ruchem oka, który dąży do zachowania jego spoczynku w pożądanej orientacji [Duc07, s. 46–47] lub tłumaczy się je jako aparat powiązany z mikrosakadami, pełniący ich rolę w przypadku ich zaniku [MCMH04]. Pomierzona amplituda drżeń i dryfów jest najczęściej mniejsza niż  $0.01^\circ$ , czyli odpowiada w przybliżeniu wielkości jednej komórki światłoczułej. Natomiast częstotliwość jest rzędu 100–150Hz [MCMH04, LZ99, s. 189–190], co czyni je niemożliwymi do zaobserwowania bez wykorzystania czułej aparatury.

Z kolei występowanie mikrosakad, odbywających się w kierunkach sprawiających wrażenie chaotycznych, jest najczęściej tłumaczone zadaniem zapobieżenia powstaniu permanentnego powidoku. Powidok ten, będący wynikiem adaptacji fotoreceptorów do nieruchomego obrazu, sprawiłby, że każdy obraz, na którym obserwator zatrzymałby się przez dłuższy czas, przestałby być widoczny [DG52, Duc07]. Innym możliwym wytłumaczeniem jest korekta położenia rzutu obrazu na siatkówce, które uległo oddaleniu od dołka środkowego w wyniku drżeń i dryfów [Cor56]. Amplituda mikrosakad wynosi poniżej trzeciej części stopnia [LZ99, s. 189–190], co odpowiada przeniesieniu obrazu o kilkaset receptorów, a czas ich trwania wynosi około 25 milisekund [MCMH04]. Należy zwrócić uwagę, że identyfikacja mikrosakad wyłącznie na podstawie amplitudy ruchu może być kłopotliwa, jako że świadome sakady również mogą odbywać się na zbliżonych odległościach.

Można się więc spodziewać, że nie jest możliwe wyodrębnienie z przebiegu orientacji gałki ocznej takich jego fragmentów, gdy pozostawałoby ono w całkowitym spoczynku.

### 1.2.2.3 Sakady i płynny pościg

Przejście pomiędzy kolejnymi fiksacjami zajmuje zwykle od 10 do 100 milisekund. Taki ruch gałek ocznych związany ze zmianą ich orientacji nazywany jest sakadą (ang. *saccade* od fr. *saccadé* oznaczającego szarpnięcie, gwałtowny ruch). Charakter tych ruchów został opisany już w drugiej połowie XIX wieku przez francuskiego

okulistę, Louis Émile Javala [Jav79]. Sakady mają dowolny zasięg w obrębie obszaru możliwego do pokrycia fiksacjami. Ich szczytowa szybkość może osiągać nawet wielkości rzędu  $500^\circ/\text{s}$  i jest zależna od amplitudy danej sakady [LZ99, s. 111]. Wykazano, że przez ten czas nie są świadomie odbierane żadne bodźce wzrokowe, można więc mówić o chwilowej ślepotce percepcji wzrokowej podczas trwania sakad [HH03].

Aparat widzenia człowieka zdaje się interpretować informacje o ruchu obiektów w sposób zupełnie odrębny, aniżeli o ich pozycji. Przykładem mającym potwierdzić tę tezę jest zjawisko, z którym można się spotkać podczas uważnego oglądania napisów kończących kinowy film, które przesuwały się ze stałą szybkością ku górze ekranu. Po pewnym czasie odnosi się wrażenie, że to nie napisy, lecz całe otoczenie znajduje się w ruchu, przesuując się nieustannie ku dołowi. Dzieje się tak wskutek adaptacji aparatu motoryczno-wizyjnego do odbieranego bodźca: pomimo konieczności wprawienia oczu w ciągły ruch, pozycja obrazu rzutowana na siatkówkę pozostaje niezmienną dając tym samym złudzenie bycia nieruchomą [LZ99, s. 193]. Podobnie można argumentować istnienie kolejnego, interesującego rodzaju ruchów gałek ocznych jakim są etapy tzw. płynnego pościgu (ang. *smooth pursuit*), które zostały po raz pierwszy odróżnione od sakad przez Dodge'a na początku XX wieku [Dod03].

Postrzeganie szczegółów obrazu jest bowiem możliwe nie tylko podczas stacjonarnych fiksacji. Ruchome bodźce wizualne, charakteryzujące się albo znacząco inną luminancją względem tła, albo kontrastem czy migotaniem, mogą być w płynny, nieprzerwany sposób śledzone przez wzrok obserwatora. Pozwala to im być rzutowanymi na tę część siatkówki, która zapewnia najlepsze widzenie. Dla dynamicznych scen i obserwacji poruszających się obiektów, płynne pościgi pełnią więc tę samą rolę, co fiksacje dla stacjonarnych bodźców. Analogicznie, dla takiego przypadku etapy pościgu będą więc rozdzielone ruchami sakadycznymi. Na przykład podczas przytoczonego już wcześniej przykładu czytania poruszającego się tekstu, odczytywanie informacji odbywa się na etapach płynnego pościgu, które stanowią logiczny odpowiednik fiksacji dla nieruchomego tekstu [VGS13].

Płynny pościg następuje z opóźnieniem rzędu 100 milisekund po wystąpieniu intencji skupienia uwagi wizualnej na ruchomym bodźcu, przy czym czas ten jest niezależny od szybkości dla obiektów poruszających się z szybkością powyżej  $5^\circ/\text{s}$ . Dla obiektów poruszających się wolniej opóźnienie ulega wydłużeniu, natomiast nie udało się zauważyć płynnych pościgów dla bardzo wolno poruszających się bodźców, których szybkość była na poziomie  $0.03^\circ/\text{s}$  i mniejsza [CG87].

W przypadku gdy ruch bodźca staje się zbyt szybki i przekracza  $30^\circ/\text{s}$  lub gdy kontrast bodźca jest zbyt niski [SKB<sup>+</sup>05], zamiast jednego, długiego płynnego pościgu za śledzonym obiektem, mamy do czynienia z sekwencją krótkich faz pościgu lub fiksacji, rozdzielonych sakadami. Takie zjawisko może występować częściej przy określonych chorobach układu nerwowego takich jak np. schizofrenia i autyzm, co jest wykorzystywane jako sposób ich diagnozy [HTA<sup>+</sup>05, TML<sup>+</sup>04, LZ99, s. 220–226].

#### 1.2.2.4 Pozostałe ruchy gałek ocznych

Innym, istotnym z punktu widzenia postrzegania głębi ruchem oczu jest konwergencja (ang. *vergence*). W przeciwieństwie do wcześniej wymienionych, istotą kon-

wergencji jest skorelowane poruszanie się obu gałek ocznych. Odbywa się to w celu zredukowania zjawiska diplopii, a więc podwójnego widzenia. Zależnie od odległości oglądanego przedmiotu, konieczna jest więc taka rotacja oczu, by osie widzenia przecięły się w miejscu skupienia uwagi. W ten sposób aparat widzenia dąży do minimalizacji rywalizacji (ang. *binocular rivalry*) dwu obrazów, pochodzących z siatkówek każdego z oczu. Względna różnica orientacji gałek ocznych może więc być przesłanką mówiącą o dystansie na jakim skupiony jest wzrok obserwatora (zob. Rozdz. 1.1.2.1).

Oprócz fiksacji, sakad, płynnego pościgu i konwergencji, do czynności wzrokowego aparatu motorycznego zalicza się także odruch przedsionkowo-oczny (ang. *vestibulo-ocular reflex*). Pozwala on kompensować ruchy rotacyjne i translacyjne głowy, zapewniając możliwie nieznaczny ich wpływ na przemieszczenie obrazu na siatkówce oka podczas fiksacji oraz płynnego pościgu. Pomimo zmiany orientacji gałki ocznej względem głowy, nie powoduje on więc zmiany obiektu który jest odbiorcą uwagi wizualnej. Ten bezwarunkowy odruch jest więc neutralny z punktu widzenia uwagi wizualnej, a służy jedynie jej utrzymaniu.

## Podsumowanie rozdziału

Chcąc osiągnąć jak największy realizm prezentowanego użytkownikowi obrazu, niezależnie od tego czy celem jest dążenie do spowodowania iluzji miejsca czy też iluzji pozorności, konieczne jest zakomunikowanie przestrzenności wizualizowanej sceny. Zanurzeniu w wirtualnym świecie sprzyja odczuwanie głębi oglądanych obrazów, więc wywołanie tego odczucia powinno bazować na jak najwierniejszym dostarczeniu obserwatorowi wskazówek dotyczących odległości pomiędzy obiektami występującymi w oglądanej wizualizacji. Wizualizacje te zwykle nie zawierają informacji prowadzących do wytworzenia wszystkich wskazówek z realnego świata. Z tego powodu konieczne może okazać się imitowanie podczas syntezy obrazu niektórych zjawisk wizualnych, wspomagając się także wiedzą o chwilowym ukierunkowaniu uwagi wizualnej, wynikającą z pomierzonych ruchów oczu użytkownika.

## Rozdział 2

# Modelowanie kierunkowości widzenia

W tym rozdziale omówiony zostanie zbiór zagadnień powiązanych z technicznymi aspektami systemów komputerowej estymacji kierunku spojrzenia, a także ich wykorzystaniem w dziedzinie interaktywnej grafiki komputerowej.

Przybliżenie zasady działania systemów okulograficznych (Rozdz. 2.1) doprowadzi do wskazania i zbadania istotnego problemu jakości danych pozyskiwanych z pomocą okulografii, będącego ważnym ograniczeniem współcześnie stosowanych metod (Rozdz. 2.2). Problem ten wymaga stosowania, dopasowanych do konkretnego zastosowania, sposobów redukcji jego wpływu na uzyskiwane rezultaty. Przedstawiona zostanie powstała w ramach prac nad niniejszą rozprawą, przeznaczona dla rozwiązań związanych z grafiką komputerową, technika identyfikacji obiektu, na którym spoczywa w danej chwili uwaga obserwatora (Rozdz. 2.3). Jej działanie zostanie poddane weryfikacji względem wyników otrzymywanych z użyciem innych, popularnych metod (Rozdz. 2.3.2). Ocena rozwiązania będzie wiązała się z wprowadzeniem dedykowanej dla tego problemu, percepcyjnej metryki jakości (Rozdz. 2.3.2.1). Treść tej części Rozdziału opiera się na pracach zrealizowanych w ramach współautorskich publikacji [MKNB12, MBM13, MB14].

W końcowej części rozdziału zostaną omówione i poddane badaniu skuteczności przykłady zastosowania proponowanej techniki w różnych obszarach powiązanych z grafiką komputerową. Poruszona zostanie kwestia, opartej na identyfikacji obiektów, analizy skupienia uwagi podczas oglądania materiałów wideo (Rozdz. 2.4.1, na podstawie autorskiego artykułu [BM14a]). Kolejnym przytoczonym zastosowaniem będzie optymalizacja ilości niezbędnych obliczeń, służących komputerowej syntezie obrazu, poprzez redukcję jego złożoności dla tych obszarów, które nie są objęte najbardziej szczegółowym widzeniem użytkownika (Rozdz. 2.4.2, opublikowano w [TB10]). Przedstawione zostaną także przykłady sposobów wykorzystania danych o elemencie sceny na którym spoczywa wzrok jako części interfejsu użytkownika w aplikacji grafiki czasu rzeczywistego (Rozdz. 2.4.3, opracowane w ramach autorskiej publikacji [BM14b]).

## 2.1 Okulografia

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 (zob. Rozdz. 1.2). 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.

### 2.1.1 Kierunek spojrzenia

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.

Należy odróżnić kierunek widzenia wyznaczony przez oś widzenia odpowiednio zorientowanej gałki ocznej, od orientacji uwagi wizualnej obserwatora. Skupienie uwagi odbywa się równoległe do postrzegania obrazu i może być od niej niezależne, a więc niezwiązane bezpośrednio z faktycznym ruchem oczu. Choć nie jest wykluczone, że może na ten ruch wpływać, jeśli na jej podstawie w świadomości powstanie przewidywanie odnośnie położenia obiektu zainteresowania [PSD80].

O ile więc kierunek widzenia w sensie orientacji gałek ocznych jest rzeczą obserwowalną i stosunkowo łatwą do zdefiniowania w fizyczny sposób, a jego pomiar jest możliwe poprzez obserwację gałki ocznej, o tyle dla uzyskania wiedzy na temat skupienia uwagi konieczny byłby wgląd w procesy percepcyjne i myślowe człowieka. Dlatego od współczesnych urządzeń technicznych bazujących na obserwacji oka w celu wyznaczenia kierunku patrzenia, można oczekiwać jedynie określenia kierunku w rozumieniu jego fizycznej wartości.

Natomiast co do kognitywnego ujęcia obiektu uwagi w takich warunkach konieczne jest poczynienie założenia, że w danej chwili jest on tożsamy albo z jednym z elementów obserwowanej sceny który znajduje się na osi wyznaczającej kierunek patrzenia, albo innym obiektem wyznaczonym na podstawie pewnego rodzaju heurystyki (zob. Rozdz. 2.3).

W przypadku uproszczonego założenia, że fizyczny kierunek patrzenia implikuje 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, słusznym 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 [DPHW11a].

### 2.1.2 Eye tracking, gaze tracking i okulografia

Angielskojęzyczne pojęcie *eye trackingu*, a więc śledzenia oczu, może okazać się dość niejednoznaczne, choć często jest przytaczane jako bezpośrednie tłumaczenie okulografii na język angielski. W podstawowym rozumieniu mianem tym można określić techniki pozwalające na wykrycie i śledzenie samego położenia oczu w przestrzeni obrazu, bez zważania na szczegółową analizę ich stanu czy też kierunku w jakim są one zwrócone. Ta definicja zdaje się być szczególnie chętnie przyjmowana w kręgach marketingowych związanych z rynkiem przenośnych urządzeń elektronicznych, gdzie za *eye tracker* zwykle się przyjmować algorytmy prostego wykrywania oczu w obrazie z kamery skierowanej na twarz użytkownika [Lea13]. Dopiero inne, znacznie węższe znaczeniowo określenie *gaze trackingu* czyli śledzenia spojrzenia, nie pozostawia wątpliwości co do oczekiwanego rezultatu takiego procesu.

Pojęcie okulografii odnosi się przede wszystkim do *eye trackerów* i obejmuje ogół technik powiązanych z rejestrowaniem położenia i ruchu oczu [Wil06]. W literaturze przedmiotu w przeważającej większości przypadków stosuje się jednak ogólne pojęcie *eye trackingu* dla metod, które mają za zadanie wyznaczyć albo kierunek spojrzenia, albo położenie punktu skupienia wzroku we współrzędnych zadanego medium wyświetlającego bodziec. Można by więc polemizować, czy słusznym jest stosowanie pojęcia okulografii w znaczeniu *gaze trackingu*. Jednak dla uniknięcia wprowadzania zbędnego zamieszania nomenklaturowego w stosunku do innych prac badawczych, w niniejszej pracy pojęcia *eye trackingu* i okulografii będą stosowane zamiennie, odnosząc się do powszechnie rozumianego śledzenia kierunku i punktu patrzenia.

### 2.1.3 Metody okulograficzne

Pierwsze obserwacje kierunku spojrzenia odbywały się w sposób naoczny. Istotnym kamieniem milowym stało się opisanie sakad i fiksacji mających miejsce naprzemiennie podczas czytania drukowanego tekstu, co było sprzeczne z wcześniejszym przekonaniem o płynnym ruchu oczu śledzącym kolejne wersy [Jav79]. Na przestrzeni lat zaczęła się pojawiać różnego rodzaju aparatura techniczna ułatwiająca precyzyjne wyznaczenie chwilowego kierunku patrzenia i na jego rejestrację, wykorzystująca podejścia takie jak znieczulenie gałki ocznej, jej stabilizację i znaczniki umieszczane bezpośrednio na jej powierzchni [HNA<sup>+</sup>11a, s. 20].

Wśród współczesnych podejść do tematyki śledzenia wzroku wyróżnić można przede wszystkim techniki takie jak: oparta na elektrodach umieszczonych na powierzchni skóry elektrookulografia (*EOG*, ang. *Electro-OculoGraphy*), wykorzystanie w polu magnetycznym specjalnie spreparowanych soczewek kontaktowych z zanurzoną w niej cewką, w której indukowany jest przepływ prądu, czy wreszcie oparta na obrazie z kamery wideookulografia [Duc07, s. 51–59]. W tym ostatnim podejściu wyznaczanie kierunku spojrzenia odbywa się na podstawie kolejnych klatek wideo obrazujących ruch oczu. Technika ta, za pioniera której uznaje się znanego psychologa, Alfreda Yarbusa [YHR67], jest na dzień dzisiejszy najpowszechniej stosowaną

i najbardziej rozwiniętą metodą. Te właśnie będą przedmiotem dalszych rozważań zawartych w niniejszej pracy.

### 2.1.3.1 Wideookulografia

Wideookulografy można podzielić na dwie grupy pod względem umiejscowienia kamery rejestrującej obraz oczu (lub kamer). Kamera bowiem może być albo umieszczona bezpośrednio przy oku i pozostawać nieruchomo względem głowy. Może też być związana z położeniem prezentującego bodźce medium, w przestrzeni którego planujemy uzyskać rezultaty.

W pierwszym przypadku zwykle dąży się do sytuacji w której oko wypełnia większą część kadru. Kamera może być wówczas skierowana bezpośrednio na oko lub na zwierciadlany element, w którym odbija się jego obraz (np. umieszczona pod kątem przezroczysta płytka, która nie blokuje pola widzenia lub przezroczyste szkło pokryte folią selektywnie odbijającą światło zależnie od długości fali). Pomimo korzyści płynącej z wysokiej rozdzielczości obrazu samego oka, rozwiązanie to obarczone jest zasadniczym problemem: jeśli kamerę umieścimy na głowie w taki sposób, że będzie się poruszała wraz z nią, będziemy mogli jedynie wyznaczyć kierunek spojrzenia względem głowy badanej osoby, a więc zupełnie ignorując wpływ orientacji samej głowy na ten kierunek w ujęciu całej przestrzeni otoczenia. W takim wypadku, aby wyznaczyć punkt skupienia wzroku w przestrzeni interesującego nas, niezwiązanego z głową medium, konieczne jest albo zapewnienie jego stabilności względem głowy poprzez unieruchomienie badanego [MKNB12], albo potraktowanie wykrywania orientacji głowy jako osobnego problemu prowadzącego do pozyskania ostatecznego wyniku [AEC96].

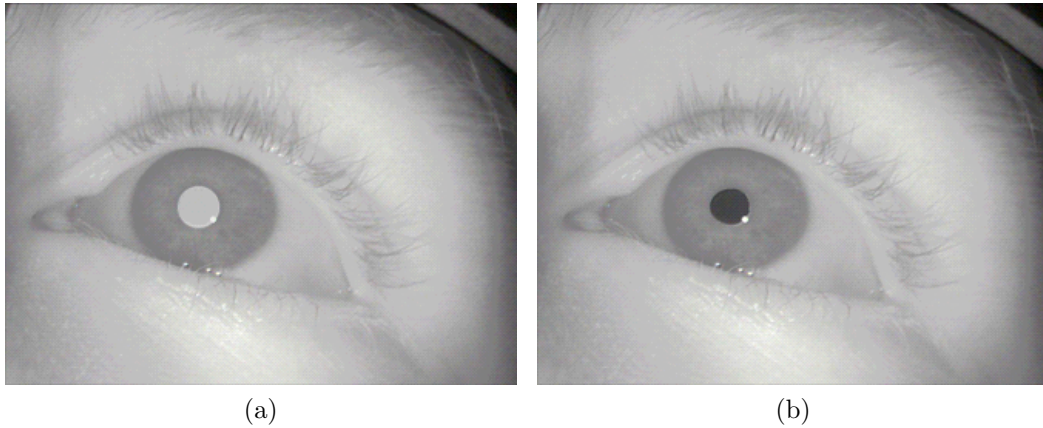
Natomiast wideookulografy, w których kamera związana jest nieruchomo z medium prezentującym bodźce i znajduje się w pewnej odległości od badanego, nazywa się eye trackerami zdalnymi (ang. *remote eye trackers*). Najczęściej można spotkać się z rozwiązaniami w których kamera umieszczona jest w sposób ściśle związany z monitorem komputerowym albo innego rodzaju wyświetlaczem, takim jak na przykład ekran telefonu komórkowego [tob13]. Obraz oczu rejestrowany ze sporej odległości (zwykle 50–70 centymetrów), obejmuje często całą głowę użytkownika wraz z otoczeniem. Można się więc spodziewać, że liczba pikseli przypadająca na każde z oczu jest tutaj znacznie mniejsza niż w przypadku eye trackerów z kamerą montowaną w bliskim sąsiedztwie oka, co przekłada się na łatwość i dokładność z jaką rozpoznawane są ich wizualne cechy służące do estymacji kierunku spojrzenia [MKNB12].

### Położenie źrenicy

Zbieranymi na bieżąco danymi, które bezpośrednio służą wyznaczeniu kierunku widzenia, są położenia środków źrenic, a także opcjonalnie położenia jednego lub więcej odbić rogówkowych w każdym z oczu. Wszystkie te dane uzyskuje się metodami przetwarzania obrazów, analizując klatki dostarczone przez kamerę wbudowaną w eye tracker.

Okulografy pracują często w oparciu o obraz obojga oczu. Jednak jeśli celem śledzenia kierunku widzenia jest określanie punktu skupienia wzroku na płaszczyźnie



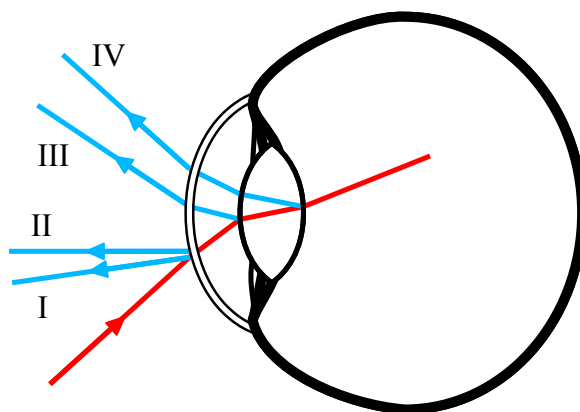


Rysunek 2.1: Dwa rodzaje obrazów oka w podczerwieni, uzyskane w uzależnieniu od wzajemnego położenia źródła światła, kamery i oka. Jeśli źródło światła znajdzie się w pobliżu osi łączącej oko z kamerą, wówczas przez otwór źrenicy będzie widoczna oświetlona część siatkówki wyścielającej dno oka (a). Jeśli zaś źródło światła znajdzie się daleko od tej osi, przez źrenicę będzie widoczna zacieniona część siatkówki (b).

bądź obiekcie, który nie będzie zmieniał odległości względem obserwatora przez cały czas trwania kalibracji i badania, to zakładając poprawne funkcjonowanie aparatu widzenia wystarczy ograniczenie się tylko do jednego oka. Wówczas wykorzystanie położenia źrenicy i odbić rogówkowych obojga oczu służy najczęściej jedynie redukcji zaszumienia danych i redukcji negatywnego wpływu na wyniki ewentualnych, chwilowych problemów z wykryciem cech jednego z oczu.

Odnalezienie źrenicy na obrazie oka zazwyczaj sprowadza się do poszukiwania jej granicy z tęczówką [MM05]. Z racji niewielkiego kontrastu jaki cechuje tę granicę w świetle widzialnym, często pozyskuje się obraz w podczerwonej części spektrum fal świetlnych. Doświetlając otoczenie oczu za pomocą światła podczerwonego, możliwe jest uzyskanie dwóch przypadków zależnych od wzajemnego położenia źródła światła, kamery i oka, które zobrazowano na rysunku 2.1. W ten sposób znacząco zwiększa się kontrast poszukiwanej granicy, co ułatwia znalezienie odpowiednio najciemniejszego bądź najjaśniejszego fragmentu obrazu. Staje się on kandydatem, by zostać uznany za źrenicę. Spotyka się także rozwiązania w których wykorzystuje się zarówno efekt jasnej, jak i ciemnej źrenicy, poprzez zsynchronizowanie dwóch lub więcej różnie rozmieszczonych źródeł światła z rejestrowaniem klatek przez kamerę i zbieranie kolejnych klatek przy naprzemiennym załączaniu każdego z nich. W ten sposób po odjęciu od siebie następujących po sobie klatek i progowaniu, niskim kosztem uzyskuje się zgrubne położenie źrenicy [ES93]. Podobny efekt można uzyskać poprzez użycie źródeł światła o różnej długości fal lub spolaryzowanych w połączeniu z kilkoma dopasowanymi elementami światłoczułymi [TIK90]. Aby opisać kształt źrenicy i znaleźć jej środek, możliwe jest też posłużenie się metodami dopasowywania elipsy [LWP05], dzięki czemu możliwe jest uzyskanie podpikselowej dokładności [ZY02].

Rysunek 2.2: Mechanizm powstawania odbić rogówkowych, czyli obrazów Purkyniego I–IV rzędu. Czerwoną linią oznaczono promień światła docierający do oka. Kolor niebieski oznacza odbicia kolejno od: zewnętrznej powierzchni rogówki (obraz I rzędu), wewnętrznej powierzchni rogówki (obraz II rzędu), zewnętrznej powierzchni soczewki (obraz III rzędu) i wewnętrznej powierzchni soczewki (obraz IV rzędu).



### Odbicie rogówkowe

Drugą z cech obrazu która często jest wykorzystywana przy estymacji kierunku spojrzenia, są odbicia rogówkowe. Na szklistej powierzchni rogówki, której kształt jest zbliżony do wycinka sfery, możliwe jest łatwe zauważenie odbić tych źródeł światła, które ją oświetlają. Rozbłyski nie pochodzą jednak wyłącznie z odbicia w zewnętrznej powierzchni rogówki, ale także jej wewnętrznej strony oraz z obu, zewnętrznej i wewnętrznej, powierzchni soczewki. Powstają więc na granicach warstw zewnętrznej struktury gałki ocznej. Rozbłyski te (ang. *glints*) nazywa się obrazami Purkyniego od nazwiska czeskiego badacza, który dokonał ich opisu (patrz: Rys. 2.2).

Gdy źródło światła pozostaje nieruchome względem kamery rejestrującej obraz oka, można zaobserwować, że obraz Purkyniego I rzędu pozostaje stacjonarnym punktem odniesienia podczas ograniczonych ruchów głową. Wykorzystanie informacji o położeniu odbicia rogówkowego może zatem posłużyć w ograniczonym stopniu do uodpornienia metody śledzenia kierunku spojrzenia na zmianę położenia i orientacji głowy badanej osoby [MM05]. Należy jednak zwrócić uwagę na to, że po przekroczeniu kąta  $\pm 12 - 15^\circ$  pomiędzy osią widzenia a osią kamery, odbicie źródła światła może pojawić się już poza sferyczną powierzchnią rogówki, tym samym zmniejszając skuteczność tej metody [SFHHFLL91].

### Inne techniki

Innym podejściem związanym ze śledzeniem odbić rogówkowych wielu źródeł światła, może być metoda opracowana przez Yoo. Wykorzystuje ona cztery źródła światła rozmieszczone na narożnikach wyświetlacza komputerowego. Po odnalezieniu odbić każdego z nich i wykorzystaniu modelu geometrii oka, możliwe jest wyznaczenie przekształcenia prowadzącego do uzyskania projekcji na powierzchni rogówki odpowiadającego zaobserwowanemu rozłożeniu. To zaś prowadzi do określenia trój-

wymiarowej orientacji gałki ocznej względem ekranu [YKLC02].

Często celem optymalizacji ilości niezbędnych obliczeń, wprowadza się pośredni krok w postaci zgrubnego wykrywania i śledzenia twarzy badanej osoby, bądź okolicy jej oczu. Dopiero w obrębie tak zawężonego obszaru dokonuje się poszukiwania źrenicy i odbić rogówkowych. Wyznaczenie części pola widzenia kamery w której znajdują się oczy (ang. *ROI, Region Of Interest*) może także służyć redukcji liczby pikseli z których składają się klatki pobierane z kamery. To z kolei skutkuje możliwością znacznego zwiększenia częstotliwości rejestrowania klatek.

## 2.1.4 Mapowanie danych okulograficznych

Znajomość położenia źrenic lub odbić rogówkowych jest pierwszym krokiem do estymacji kierunku spojrzenia lub położenia punktu skupienia wzroku. Wyznaczenie tych ostatecznych wartości wymaga zamapowania znanych położzeń, określonych w przestrzeni obrazu z kamery skierowanej na oko bądź oczy, w przestrzeń w której określone mają być dane wyjściowe. Na przykład może być to przestrzeń ekranu na którym prezentowany jest bodziec lub przestrzeń kamery zapisującej obraz sceny znajdującej się przed obserwatorem. Obie z tych przestrzeni najczęściej są dwuwymiarowe.

### 2.1.4.1 Procedura kalibracyjna

Mapowanie jest zadaniem identyfikacji modelu matematycznego o zadanych współczynnikach. Współczynniki te powinny odpowiadać konkretnym warunkom, w jakich funkcjonuje dane środowisko badań: począwszy od cech samego urządzenia, geometrycznych własności jego relacji z użytkownikiem, jak i samego użytkownika. Z racji różnorodności cech anatomicznych oraz fizjonomicznych, jak i różnic w zachowaniu poszczególnych osób, niezbędne jest wyznaczenie współczynników kalibracji dla każdego użytkownika osobno. Należy więc traktować kalibrację jako procedurę indywidualną, konieczną do powtórzenia, gdy zachodzi zmiana konfiguracji przestrzennej systemu.

Zbieranie danych okulograficznych musi więc zwykle zostać poprzedzone kalibracją systemu, podczas której dokona się wyliczenia współczynników mapowania na podstawie znanych danych. Te znane dane są zbiorem odwzorowań punktów referencyjnych, który najczęściej tworzony jest na podstawie procedury kalibracyjnej polegającej za zaprezentowaniu użytkownikowi wizualizacji punktów o znanych współrzędnych w docelowej przestrzeni, poleceniu mu wpatrywania się w nie i zebraniu próbek zawierających dane o położeniu środka źrenicy i ewentualnie odbić rogówkowych.

Zazwyczaj podczas zbierania danych kalibracyjnych, użytkownikowi prezentuje się kolejne punkty kalibracyjne z siatki, która powinna pokrywać jak największą część docelowego obszaru roboczego. W ten sposób mapowanie sprowadza się do rozwiązania pary zadań dwuwymiarowej interpolacji, gdzie dwuwymiarowym wejściem są współrzędne  $x$  i  $y$  pochodzące z próbki uczącej, a jednowymiarowym wyjściem jest jedna z dwu współrzędnych kierunku spojrzenia lub położenia punktu skupienia wzroku. Poza obszarem wyznaczonym przez rozmieszczenie punktów kalibracyjnych zadanie staje się, najczęściej mniej dokładną, ekstrapolacją.

Kolejne punkty kalibracyjne prezentuje się użytkownikowi w postaci okręgów na jednolitym tle. Ekspozycja każdego z nich powinna trwać dostatecznie długo, by każda osoba była w stanie zauważyć nowe położenie, przenieść wzrok i skupić się na nowym punkcie. Aby to ułatwić, sugeruje się by położenie wyświetlanego okręgu było animowane i płynnie przemieszczało się w krótkim czasie pomiędzy docelowymi położeniami. Warto także rozważyć stopniowe, płynne zmniejszanie się promienia takiego okrągłego markera wraz ze zbieraniem kolejnych próbek dla danego punktu kalibracyjnego. Ruch obserwowanego obrazu sprzyja skupieniu wzroku, tym samym prowadząc do pozyskania bardziej stabilnych współrzędnych z okulografu.

Ze zbioru zebranych próbek należy usunąć te, które znacząco odbiegają od reszty dla danego punktu kalibracyjnego. Można tu zastosować np. metodę usunięcia kilku procent próbek najbardziej oddalonych od wartości średniej lub mediany dla danego punktu. Warto również usunąć ze zbioru te próbki, które zebrano bezpośrednio po przejściu do pokazywania punktu w nowym położeniu. W ten sposób zredukuje się możliwość wystąpienia sytuacji, w której obserwator nie zdążył jeszcze przenieść wzroku w nowe miejsce. Podobnie należy traktować próbki zebrane po wystąpieniu wadliwych pomiarów, związanych np. z zamknięciem powiek przez użytkownika. Dobrą praktyką jest wprowadzenie opóźnienia rzędu kilkuset milisekund po wykryciu takiego zdarzenia, przez który to czas próbki nie będą zbierane.

Aby wyeliminować możliwość predykcji kolejnych położenia punktów kalibracyjnych przez doświadczonych obserwatorów, co mogłoby skutkować prowadzącym do zaburzenia jakości zebranych próbek przedwczesnym przeniesieniem punktu skupienia wzroku, warto też rozważyć przechodzenie do kolejnych punktów kalibracyjnych w sposób losowy. Przy czym dobrą praktyką w celu przygotowania użytkownika do skupienia się na procedurze, jest rozpoczęcie kalibracji od wyświetlenia punktu znajdującego się w centrum obszaru roboczego. Podczas ekspozycji tego początkowego punktu dane nie będą zbierane, musi więc zostać ponownie wyświetlony później, dołączając do zbioru pozostałych punktów.

Przejścia do kolejnych punktów mogą odbywać się albo w sposób zautomatyzowany po zebraniu odpowiedniej, spodziewanej liczby poprawnych próbek, albo w sposób manualny przez np. wciśnięcie odpowiedniego klawisza. Ta druga metoda jest jednak bardziej kłopotliwa i jej użycie przy niedoświadczonych obserwatorach wymaga ingerencji operatora. Dlatego zautomatyzowana procedura kalibracyjna zdaje się być trafniejszym wyborem dla popularnych zastosowań okulografii.

## Obliczenie współczynników

Ujęcie odwzorowań przestrzeni kamery w przestrzeń docelową jako zbiór liniowo niezależnych równań, gdzie zwykle każdemu punktowi kalibracyjnemu o znanych współrzędnych ekranowych odpowiada więcej niż jedna próbka ucząca, sprowadza kwestię obliczenia współczynników mapujących do problemu rozwiązania nadokreślonego układu równań. Wylczenie współczynników modelu mapującego jest więc zadaniem minimalizacyjnym, gdzie dąży się do uzyskania jak najmniejszego błędu, godząc się na brak jednoznacznego rozwiązania. Poniżej przedstawiono propozycję znalezienia takiego pseudorozwiązania w oparciu o metodę najmniejszych kwadratów z wykorzystaniem rozkładu macierzy równań według wartości osobliwych (ang. *SVD*, *Singular Value Decomposition*), natomiast dalej w Rozdziale 2.1.4.2 zapro-

ponowano proste użycie macierzy odwrotnej. Przy czym nie są to jedyne możliwe sposoby rozwiązania tego problemu minimalizacji.

Rozpatrzmy sytuację, w której zadaniem okulograficznym jest wnioskowanie o położeniu punktu skupienia wzroku w przestrzeni ekranu, jedynie na podstawie położenia środka źrenicy w przestrzeni kamery eye trackera. Przyjęto, że mapowanie zostanie oparte o model drugiego rzędu następującej postaci:

$$\begin{aligned}x_t &= a_0 + a_1x_p^2 + a_2x_p + a_3x_py_p + a_4y_p + a_5y_p^2, \\y_t &= b_0 + b_1x_p^2 + b_2x_p + b_3x_py_p + b_4y_p + b_5y_p^2,\end{aligned}$$

gdzie  $x_t$  oraz  $y_t$  oznaczają współrzędne wyświetlanych punktów,  $x_p$  oraz  $y_p$  to współrzędne środka źrenicy w przestrzeni kamery dostarczone przez okulograf, a  $a_i$  oraz  $b_i$  to nieznanne parametry modelu. Równania dla  $x$  oraz  $y$  można traktować oddzielnie. Kwestia doboru modelu i porównanie wyników otrzymywanych z użyciem różnych z nich zostanie poruszona dalej w tym rozdziale.

Dane ze zbioru uczącego stanowią ograniczenia rozpatrywanego modelu. Gdyby danymi wejściowymi miały być położenia środków źrenic względem odbić rogówkowych, obliczenia takich położenia należałoby dokonać przed rozpoczęciem obliczania współczynników i wyniki wykorzystać jako  $x_p$  oraz  $y_p$  w opisanej wyżej metodzie. W celu zredukowania liczby ograniczeń do jednego dla każdego wyświetlanego punktu, jak i w celu odrzucenia próbek najbardziej odległych, tzw. *outlierów*, dla każdego z  $t$  punktów można obliczyć np. medianę  $\bar{x}_p$  ze zbioru odpowiadających im próbek. Tak uzyskane dane można uznać za ograniczenia poszukiwanego modelu. W wyniku tej operacji, otrzymujemy dwa modele z 6 parametrami i  $t$  ograniczeniami każdy. Należy więc zapewnić spełnienie warunku  $t \geq 6$ , aby można było dopasować rozwiązanie problemu.

Do znalezienia wartości parametrów układu zastosowano minimalizację metodą najmniejszych kwadratów [HZ03]. Nasz przypadek można przedstawić w postaci (rozważania dla  $x_t$ , dla  $y_t$  analogicznie):

$$PA = T_x,$$

gdzie  $P$  jest macierzą wartości  $\overline{x_p^j y_p^k}$   $j, k = 0..2$ ,  $A$  poszukiwanym wektorem parametrów  $a_i$  ( $i = 0..5$ ), a  $T_x$  wektorem wartości  $x_t$ . Rozwiązaniem tego równania jest więc:

$$A = P^+T_x,$$

gdzie  $P^+$  jest macierzą pseudoodwrotną do  $P$ , określoną jako macierz, która dla elementów równych 0 przyjmuje wartość 0, a dla elementów różnych od zera - odpowiadające wartości z macierzy odwrotnej. Dla zachowania pewności bezbłędnego wykonania na maszynie komputerowej operującej na liczbach zmiennoprzecinkowych, należy dla warunku równości przyjąć tolerancję o wąskim zakresie  $\epsilon$ .

Stosując dla macierzy  $P$  rozkład *SVD* (ang. *Singular Value Decomposition*) [GK65], można przedstawić ją jako:

$$P = UDV^T,$$

co, przy uwzględnieniu własności  $U$ ,  $D$ ,  $V$  wynikających z rozkładu SVD (macierze  $U$  i  $V$  są unitarne - ich odwrotności są transpozycjami ich samych, a macierz  $D$  jest macierzą diagonalną), pozwala uprościć zadanie znalezienia macierzy pseudoodwrotnej:

$$P^+ = VD^+U^T,$$

gdzie  $D^+$  jest macierzą diagonalną, która na głównej przekątnej posiada odwrotności wartości odpowiadających jej elementów z macierzy  $D$ , z zachowaniem wartości zerowych.

Po wykonaniu finalnej operacji

$$A = VD^+U^T T_x,$$

otrzymuje się wektor wartości parametrów  $a_i$  modelu, odpowiadających warunkom narzuconym przez wartości zebranych próbek.

W rezultacie możemy dokonywać mapowania dowolnej próbki  $(x_e, y_e)$  pochodzącej z pomiaru okulograficznego, na współrzędne ekranowe  $(x_m, y_m)$ :

$$\begin{aligned} x_m &= a_0 + a_1x_e^2 + a_2x_e + a_3x_ey_e + a_4y_e + a_5y_e^2, \\ y_m &= b_0 + b_1x_e^2 + b_2x_e + b_3x_ey_e + b_4y_e + b_5y_e^2. \end{aligned}$$

#### 2.1.4.2 Dobór modelu matematycznego

Warto rozpatrzyć problem doboru modelu matematycznego stosowanego na potrzeby mapowania położenia cech oczu w przestrzeni kamery na docelową przestrzeń, w której oczekuje się określenia kierunku spojrzenia lub położenia punktu skupienia wzroku. Poniżej przedstawiono rozważania dotyczące porównania kilku prostych modeli pierwszego i drugiego rzędu oraz wyniki porównania rezultatów uzyskanych za ich pomocą. Rozważania te dotyczą danych pochodzących z urządzeń służących okulografii, nie są więc próbą uniwersalnego rozwiązania problemu mapowania zachodzącego pomiędzy takimi przestrzeniami, a jedynie praktycznym podejściem mającym za zadanie wybrać najlepszą z możliwości w istniejących warunkach i dla rzeczywistych danych wejściowych.

Dla dalszych zastosowań będziemy wykorzystywać jako dane wejściowe dwa rodzaje pozycji uzyskanych na podstawie analizy obrazu z kamery skierowanej na oczy użytkownika: pozycję w przestrzeni kamery środków źrenic, odpowiednio lewej ( $P_L$ ) i prawej ( $P_R$ ):

$$\begin{aligned} P_L &= \{x_{p_L}, y_{p_L}\}, \\ P_R &= \{x_{p_R}, y_{p_R}\} \end{aligned}$$

oraz pozycję w przestrzeni kamery obrazów Purkyniego I-go rzędu, odpowiednio dla oka lewego ( $CR_L$ ) i prawego ( $CR_R$ ):

$$CR_L = \{x_{cr_L}, y_{cr_L}\},$$

$$CR_R = \{x_{cr_R}, y_{cr_R}\},$$

a właściwymi wartościami które będziemy poddawać mapowaniu, będą różnice powyższych, czyli odpowiednio  $E_L$  oraz  $E_R$ :

$$E_L = P_L - CR_L = \{x_{e_L}, y_{e_L}\},$$

$$E_R = P_R - CR_R = \{x_{e_R}, y_{e_R}\}.$$

Na potrzeby utworzenia zbioru uczącego służącego do porównania wyników uzyskiwanych z użyciem badanych modeli, obliczana jest mediana współrzędnych  $x$  i  $y$  próbek zebranych dla każdego z punktów kalibracyjnych. W ten sposób dla każdego z ekranowych punktów kalibracyjnych uzyskiwany jest dokładnie jeden odpowiadający mu punkt określony w przestrzeni kamery, przy czym zastosowanie mediany pozwala na proste odrzucenie wyników skrajnych (ang. *outliers*), mogących być wynikiem zaszumienia lub zniekształcenia pomiarów. To właśnie wektor tych wartości jest przekazywany do algorytmu obliczającego współczynniki mapowania jako wektor danych uczących. Współrzędne ekranowe punktów kalibracyjnych są znane, można więc dokonać mapowania danych pochodzących z eye trackera, w przestrzeń ekranu.

Zatem celem badania jest wybór modelu  $F$ , który na podstawie wyliczonych współczynników kalibracji  $C$  i chwilowej próbki  $E$ , zawierającej różnicę położenia środka źrenicy i odbicia rogówkowego, pozwalałby najtrafniej wyznaczyć ekranowe współrzędne  $S$  punktu na który w danym momencie spogląda użytkownik:

$$S = F(C, E).$$

W przypadku śledzenia obojga oczu będziemy mieli do czynienia z czterema różnymi mapowaniami, a więc czterema zestawami współczynników, odpowiednio dla  $x$  i  $y$  oka lewego, jak i obu tych współrzędnych dla oka prawego. Sposób  $M$  obliczenia współczynników powinien pozwalać na ich uzyskanie na podstawie wektora danych uczących  $T_E$  oraz wektora znanych współrzędnych ekranowych wyświetlanych punktów  $W$ . Wektor danych uczących zawiera wspomniane wcześniej mediany wartości  $E$  dla kolejnych punktów kalibracyjnych:

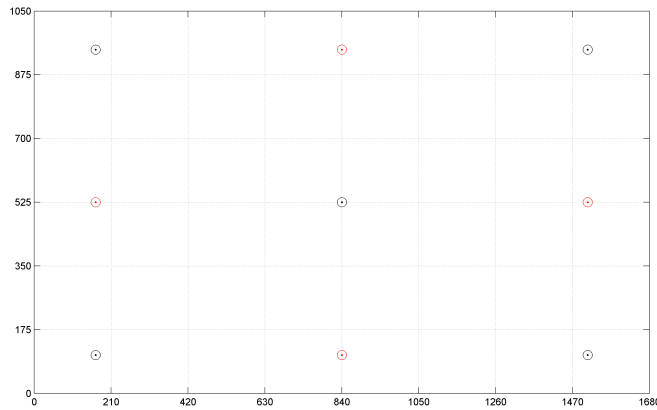
$$C_{x_L} = M(T_{E_L}, W),$$

$$C_{y_L} = M(T_{E_L}, W),$$

$$C_{x_R} = M(T_{E_R}, W),$$

$$C_{y_R} = M(T_{E_R}, W).$$

Aby pokryć całą powierzchnię roboczą ekranu oraz umożliwić określenie przebiegu płaszczyzny obrazującej mapowanie, liczba punktów kalibracyjnych dla omawianego przypadku została ustalona na pięć: cztery narożniki ekranu oraz jego środek.



Rysunek 2.3: Użyte punkty kalibracji (czarne) i punkty testowe (czerwone) ukazane na płaszczyźnie ekranu.

Punkty kontrolne dla których również zbierano dane, umieszczone były na środkach krawędzi ekranu (Rys. 2.3).

### Zbieranie danych

Każdy z punktów kalibracyjnych i kontrolnych był eksponowany przez dwie sekundy. W tym czasie obserwator był proszony o nieruchome spoglądanie w jego kierunku. Dane o położeniu środków źrenicy oraz odbić rogówkowych były rejestrowane z częstotliwością  $250Hz$ , dzięki czemu dla każdego z punktów otrzymano ok. 500 próbek. Z każdego zestawu próbek przed jakąkolwiek dalszą analizą odfiltrowano 5% tzw. *outlierów*, próbek znajdujących się w największej odległości od pozostałych [tob11].

Dane zbierane były z użyciem urządzenia *SensoMotoric Instruments RED250*. Obraz wyświetlany był na 22" monitorze w rozdzielczości  $1680 \times 1050$  pikseli. W badaniu wykorzystano dane zebrane od 12 osób (2 kobiety i 10 mężczyzn, wiek 21-42 lat).

### Modele

Przebadano cztery modele o następujących strukturach:

$$F_1 : s = c_0 + c_1x_e + c_2y_e,$$

$$F_2 : s = c_0 + c_1x_e + c_2y_e + c_3x_ey_e,$$

$$F_3 : s = c_0 + c_1x_e + c_2y_e + c_3x_e^2 + c_4y_e^2,$$

$$F_4 : s = c_0 + c_1x_e + c_2y_e + c_3x_ey_e + c_4x_e^2 + c_5y_e^2,$$

gdzie współczynniki  $c_i$  można przedstawić w postaci wektora:

$$C = \{c_0, c_1, \dots, c_n\}.$$



Identyfikacji dokonano wykorzystując prostą metodę najmniejszych kwadratów, uwzględniającą obliczenie macierzy odwrotnej. Wartości współczynników  $C$  obliczono z użyciem macierzy  $Z_{T_E}$  zawierającej zależne od danego modelu potęgi  $x_e$  oraz  $y_e$  oraz wektora znanych współrzędnych ekranowych punktów kalibracyjnych  $W$ . Zarówno dla lewego, jak i dla prawego oka mamy więc:

$$\begin{aligned} W_x &= C_x Z_{T_E}, \\ W_y &= C_y Z_{T_E}, \end{aligned}$$

co po przekształceniu pozwala obliczyć potrzebne wartości  $C$ :

$$\begin{aligned} C_x &= W_x Z_{T_E}^T (Z_{T_E} Z_{T_E}^T)^{-1}, \\ C_y &= W_y Z_{T_E}^T (Z_{T_E} Z_{T_E}^T)^{-1}. \end{aligned}$$

## Rezultaty

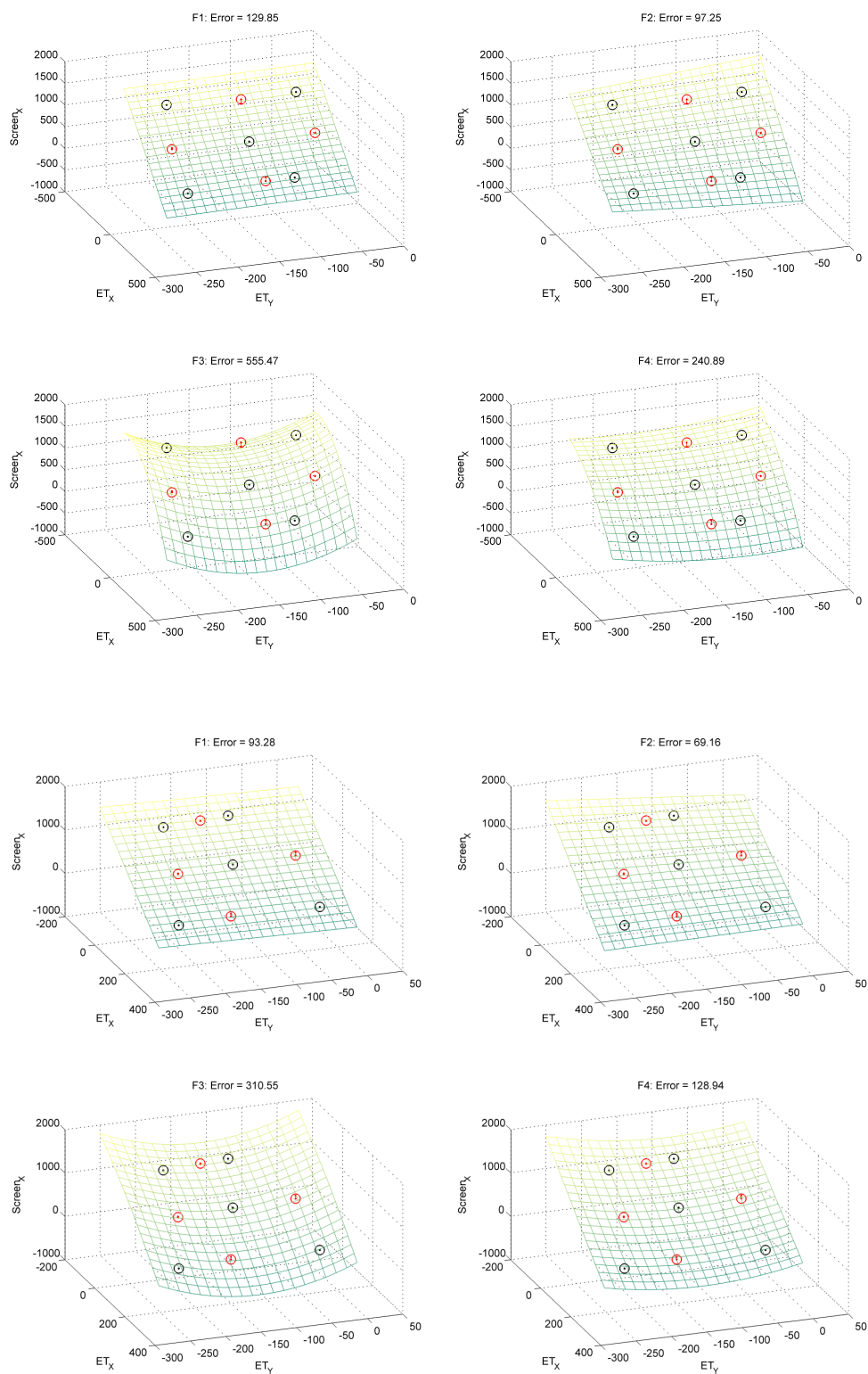
W celu dokonania oceny skuteczności poszczególnych modeli, posłużono się miarą opartą o średniokwadratowy błąd wartości otrzymywanych we wszystkich punktach referencyjnych (kalibracyjnych oraz testowych). Oddzielnie obliczono wartości dla każdego z oczu oraz każdej ze współrzędnych. Następnie uśredniono wyniki dla wszystkich obserwatorów. Porównanie przedstawiono w Tabeli 2.1.

Model	Oko lewe				Oko prawe			
	x		y		x		y	
	śr. bł.	od. std.	śr. bł.	od. std.	śr. bł.	od. std.	śr. bł.	od. std.
$F_1$	93.06	34.06	137.27	120.59	140.10	220.68	157.30	62.02
$F_2$	87.19	29.33	130.99	95.23	122.03	180.45	148.17	135.66
$F_3$	561.76	751.40	752.60	665.19	772.80	956.69	1181.31	1297.16
$F_4$	517.42	633.81	434.07	347.90	982.86	1091.01	608.99	921.38

Tabela 2.1: Wartości błędu średniokwadratowego dla testowanych modeli.

Powierzchnie wizualizujące obliczone mapowanie dla współrzędnej  $x$  oka lewego i prawego przedstawiono na Rys. 2.4. Można zaobserwować charakterystyczne, symetryczne przesunięcie na osi  $x$  jednego z zamapowanych w przestrzeń  $S$  punktów referencyjnych, wynikające prawdopodobnie z cech anatomicznych oka (przysłonięcie części źrenicy przy skrajnym kącie spojrzenia, co zaburza położenie wyznaczonego środka jej widocznej części), co zdaje się dodatkowo podkreślać nieliniowość rozpatrywanego problemu.

Na podstawie otrzymanych wyników można wysunąć wniosek, że model o strukturze  $F_2$  jest tym, który pozwala na najlepsze odwzorowanie rozpatrywanego problemu. Jego użycie poskutkowało uzyskaniem najmniejszych błędów dla warunków eksperymentu. Drugim pod względem jakości jest prosty model liniowy, jednak z racji niedostosowania do charakteru problemu, wyniki uzyskiwane z jego użyciem są



Rysunek 2.4: Porównanie modeli dla współrzędnej  $x$  lewego (pierwsze cztery diagramy) oraz prawego (cztery ostatnie diagramy) oka wybranego obserwatora. Wykreślono powierzchnie odpowiadające wyliczonym współczynnikom rozpatrywanych modeli.



Rysunek 2.5: System *DIY* podczas pracy. Konieczne jest wykorzystanie podbródka stabilizującego położenie i orientację głowy.

zauważalnie gorsze. Z kolei bardziej złożone modele wyższych rzędów nie pozwalają na uzyskanie satysfakcjonujących wyników przy tak małej liczbie punktów kalibracyjnych. Ich użycie wymagałoby więc zwiększenia liczby punktów. Podnosząc rząd wielomianu należy jednak mieć na względzie ryzyko zaistnienia zjawiska Rungego, jak i nadmiernego dopasowania mapowania do siatki danych referencyjnych.

### 2.1.5 Budowa prostego okulografu

Do niedawna istotnym problemem w dostępności gotowych rozwiązań okulograficznych była ich cena rynkowa. Koszt systemów eyetrackingowych oferowanych przez wiodących producentów w branży potrafił przekroczyć nawet 20 tysięcy euro [Sen09, Tob09]. Przy czym należy zwrócić uwagę, że znaczący udział w tej cenie przypada na wartość badań prowadzących do powstania produktu, a nie samych technicznych elementów służących do budowy urządzenia. Z tego powodu podjęto próbę budowy prostego systemu okulograficznego opartego na łatwo dostępnych komponentach, którego łączny koszt budowy nie powinien przekroczyć kwoty 50 euro. W ten sposób zaproponowano, opracowany przy współudziale autora niniejszej rozprawy, system nazwany *DIY* (ang. *Do It Yourself*) [MKNB12].

System *DIY* jest implementacją idei jednoocznego eye trackera mobilnego, a więc noszonego przez użytkownika na głowie, ale przeznaczonego do użytkowania z ekranem komputerowym. Z racji umieszczenia kamery w pobliżu oka, możliwe jest użycie kamery o niewygodnych własnościach technicznych. W celu ograniczenia kosztów problem ruchów głowy względem ekranu został całkowicie pominięty, ograniczając użytkowanie tylko do takich sytuacji, w których orientacja i położenie głowy badanej osoby nie zmieniają się. Tym samym konieczne jest użycie fizycznego stabilizatora głowy w postaci np. podbródka (ang. *chinrest*, zob. Rys. 2.5).

Porównanie jakości danych dostarczanych przez system okulograficzny *DIY* z rozwiązaniem dostępnym na rynku przedstawiono w Rozdziale 2.2. Poniżej natomiast opisano budowę i sposób działania proponowanego rozwiązania.



Rysunek 2.6: Urządzenie *DIY* zostało zbudowane na bazie okularów ochronnych oraz elementów dostępnej na rynku kamery internetowej.

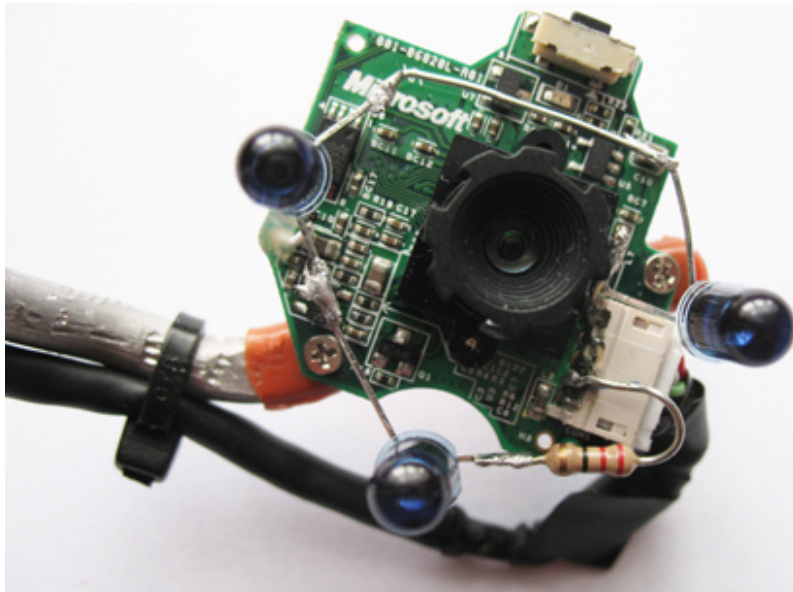
Urządzenie *DIY* składa się z dwóch podstawowych części (zob. Rys. 2.6): ramy, którą stanowi zmodyfikowana oprawa okularów ochronnych oraz modułu obrazującego, który jest do niej przymocowany.

#### 2.1.5.1 Moduł obrazujący

Podstawą modułu obrazującego jest kamera służąca do rejestrowania obrazu oka użytkownika z niewielkiej odległości (ok. 5 cm w przypadku omawianego rozwiązania). W celu ograniczenia kosztów i użycia elementów, które są powszechnie dostępne, omawiane urządzenie oparto na kamerze internetowej *Microsoft Lifecam VX-1000* (zob. Rys. 2.7). Po usunięciu jej obudowy dokonano jednak kilku modyfikacji.

Kamera *VX-1000* jest w stanie dostarczać poprzez interfejs *USB 2.0* obraz z częstotliwością 30 klatek na sekundę, obraz rejestrowany jest przez sensor *CMOS* o rozdzielczości  $640 \times 480$  pikseli (0.31 megapikseli). Taki rozmiar obrazu jest wystarczający do precyzyjnego odnalezienia położenia środka źrenicy, gdy obraz oka z założenia wypełnia dużą część kadru, co ma miejsce w przypadku tak bliskiego umieszczenia kamery przy kącie widzenia obiektywu zbliżonym do  $55^\circ$  (por. Rys. 2.8). Przetwornik kamery swoją czułością pokrywa oprócz światła widzialnego także bliską podczerwień, jednak ta została zablokowana poprzez wbudowany w obiektyw filtr podczerwieni. Jako że celem wykorzystania kamery jest praca w tej części spektrum światła, filtr ten został usunięty i zastąpiono go filtrem światła widzialnego przepuszczającym podczerwień. Tenże, dostatecznej jakości, uzyskano z wywołanego filmu fotograficznego.

Aby doświetlić oko i uzyskać wysoki kontrast na granicy źrenicy i tęczówki, który pozwoli na łatwe wykrycie położenia źrenicy, moduł obrazujący wyposażono w diody *LED* emitujące światło podczerwone. W celu uzyskania równomiernego oświetlenia i uniknięcia nadmiernych rozbłysków w powierzchni rogówki, użyto trzech diod otaczających obiektyw kamery. Diody, połączone szeregowo z rezystorem  $22 \Omega$ , zasilane



Rysunek 2.7: Moduł obrazujący urządzenia *DIY*. Na aluminiowym pręcie zamontowano pozbawioną obudowy, zmodyfikowaną kamerę internetową *Microsoft Lifecam VX-1000*.



Rysunek 2.8: Przykładowy obraz zarejestrowany przez kamerę skierowaną na oko użytkownika eye trackera *DIY*. Oko wypełnia dużą część kadru, przez co możliwe jest precyzyjne znalezienie środka źrenicy. Widoczne są trzy odbicia rogówkowe podczerwonych diod doświetlających oko. Przez źrenicę widoczny jest zaciemniony obszar siatkówki, przez co mamy do czynienia ze zjawiskiem *dark pupil* ułatwiającym odnalezienie źrenicy.

są napięciem 5 V z tego samego źródła, co kamera: z przewodu *USB*.

Każdą z zastosowanych w module obrazującym diod emitujących światło podczerwone, cechuje moc emisji na szczytowym poziomie nie większym niż  $150 \text{ mW sr}^{-1}$  dla długości fali 940 nm. Jest to wartość znacząco niższa od progu ekspozycji określonego przez wytyczne ICNIRP (*International Commission on Non-Ionizing Radiation Protection*) [oNIRP+13]. Korzystanie z bliskiego doświetlenia oka podczerwienią w takiej konfiguracji jest więc całkowicie bezpieczne nawet podczas długotrwałego użytkowania.

### 2.1.5.2 Oprogramowanie

Zadanie modułu obrazującego kończy się na dostarczeniu systemowi komputerowemu kolejnych klatek obrazu z kamery. Przekształcenie tego obrazu w możliwe do wykorzystania dane o kierunku spojrzenia i punkcie skupienia wzroku są realizowane programowo.

W omawianych zastosowaniach posłużono się udostępnionym na licencji *open source* oprogramowaniem *ITU Gaze Tracker*, opracowanym na uniwersytecie w Kopenhadze [itu12]. Odpowiada ono za przeprowadzenie procedury kalibracji, wykrycie położenia środka źrenicy oraz dokonanie jego mapowania na płaszczyznę ekranu w celu określenia położenia punktu skupienia wzroku.

Wyjście z oprogramowania *ITU* konsumowane jest przez jeden z autorskich interfejsów programowych (biblioteka napisana w języku *C++*, biblioteka *.NET*, interfejs *MATLAB* oparty o technikę *MEX*), pozwalających na dalsze wykorzystanie danych w końcowych aplikacjach. Ostatecznie dane dotyczące punktu skupienia wzroku mogą być wykorzystane w dowolny sposób, zarówno w charakterze *on-line* (na bieżąco, w czasie zbliżonym do rzeczywistego), jak i *off-line* (analiza danych po ich zebraniu). System *DIY* sprawdził się i był wykorzystywany przy wszystkich zastosowaniach okulografii opisanych dalej w niniejszej rozprawie.

### 2.1.5.3 Dalszy rozwój

Naturalnym dalszym etapem rozwoju systemu *DIY* jest próba wyeliminowania problemu niemożności zmiany orientacji i pozycji głowy podczas korzystania z okulografu. Chcąc pozostać przy koncepcji urządzenia noszonego na głowie, konieczne będzie posłużenie się pewną techniką śledzenia głowy (ang. *head tracking*), pozwalającą na wyznaczenie względnej relacji pomiędzy urządzeniem (czyli *de facto* głową), a ekranem na płaszczyznę którego rzutowane są wyniki. Odwrotnym podejściem prowadzącym do tego samego celu mogłoby być wykrywanie czworokąta ekranu w obrazie pochodzącym z dodatkowej kamery zamontowanej na urządzeniu, ale skierowanej w stronę sceny oglądanej przez użytkownika.

Umieszczenie na urządzeniu dodatkowej kamery pozwalającej na uzyskanie obrazu świata znajdującego się przed użytkownikiem mogłoby dodatkowo poszerzyć spektrum zastosowań systemu *DIY*. Poprzez rzutowanie kierunku spojrzenia na obraz świata możliwe byłoby użycie okulografu do badań terenowych, gdzie bodźcami nie są obrazy prezentowane na ekranie, a obiekty pochodzące ze świata rzeczywistego. Taka koncepcja wiąże się jednak z zaistnieniem problemów natury geometrycznej

(np. zróżnicowanie odległości płaszczyzn na których spoczywa punkt skupienia wzroku i idące za tym problemy kalibracyjne) czy też powiązanych z poruszaniem się systemu okulograficznego (np. konieczność klasyfikacji ruchów oczu z uwzględnieniem zmiany orientacji głowy).

Innym kierunkiem rozwoju mogłaby być poprawa jakości danych. Użycie lepszej kamery, próbkowania z większą częstotliwością i innych, szybszych i dokładniejszych algorytmów wykrywania cech obrazu oka [Wol15], prowadzi do zmniejszenia błędów śledzenia i zwiększenia szeroko rozumianej efektywności działania systemu.

## 2.2 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.

### 2.2.1 Czynniki wpływające na jakość

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 (por. Rozdz. 2.1.4.2).

#### 2.2.1.1 Techniczne możliwości kamery

Podstawą do wyznaczenia kierunku spojrzenia obserwatora jest w przypadku wideo-okulografii obraz oka zarejestrowany przez kamerę. Cechami, które muszą zostać wyodrębnione z tegoż obrazu, a które mają znaczenie dla procesu estymacji kierunku spojrzenia, są najczęściej położenia środka źrenicy oraz odbić rogówkowych. Oczywistym jest, że obraz w niskiej rozdzielczości utrudnia precyzyjne określenie tych położenia, szczególnie uwzględniając zwykle niedoskonałe warunki oświetleniowe i zmienność rejestrowanego obrazu wywołaną ruchem użytkownika bądź samej kamery. Należy tutaj zwrócić uwagę, że w eye trackerach, w których kamera zamontowana jest w sposób związany z położeniem głowy i skierowana jest na oko, najczęściej zadowalające rezultaty można uzyskać już z kamerami, które dostarczają obraz o stosunkowo niewielkiej rozdzielczości (np. 640 na 480 pikseli jak w przykładzie opisanym w Rozdz. 2.1.5), ponieważ obraz oka wypełnia znaczną część każdej klatki. Gdy jednak śledzenie wzroku odbywa się w sposób zdalny, z dopuszczeniem ograniczonego ruchu głowy, obraz oka stanowi jedynie niewielką część kolejnych klatek. Tym samym podejście to wymusza stosowanie kamer o podwyższonej rozdzielczości,



by zapewnić wystarczającą zdolność do rozróżnienia kształtów i położeń poszukiwanych cech obrazu.

Liczba rejestrowanych klatek w jednostce czasu również jest istotnym czynnikiem dla procesu okulografii. Z uwagi na bardzo krótki czas trwania ruchów sakadycznych (por. Rozdział 1.2.2.3), jeśli celem śledzenia kierunku spojrzenia ma być wyodrębnienie charakterystyki tychże, kamera winna być w stanie pracować z częstotliwością większą niż  $100\text{Hz}$  (czas trwania ruchów sakadycznych może wynosić 10 milisekund, por. Rozdz. 1.2.2.3). Na dzisiejszym rynku dominują jednak rozwiązania pracujące z częstotliwością w zakresie  $50 - 60\text{Hz}$ , co zdaje się być wystarczające dla podstawowych zastosowań, gdzie oczekiwanym rezultatem jest informacja o występowaniu kolejnych fiksacji obserwatora. Należy także wziąć pod uwagę możliwość wystąpienia zjawiska rozmycia ruchowego (ang. *motion blur*), jeśli czas naświetlania kolejnych klatek będzie dostatecznie długi, co wpływa niekorzystnie na zdolność do określenia położeń poszukiwanych elementów obrazu. Inne cechy typowe dla kamer, jak na przykład zaszumienie obrazu wynikające z własności elementu światłoczułego lub optyki, również mają istotne znaczenie dla jakości uzyskiwanych rezultatów i mogą być źródłem przekłamań w zebranych i obliczonych danych.

### 2.2.1.2 Ograniczenia stosowanych modeli

Wyznaczanie kierunku spojrzenia na podstawie obrazu oka odbywa się poprzez podanie na wejście obranego matematycznego modelu położeń środka źrenicy i ewentualnie położeń odbić rogówkowych. Model taki, zasilony uprzednio danymi zebranymi podczas procedury kalibracyjnej dla konkretnego użytkownika, pozwala na wyznaczenie współrzędnych kierunku spojrzenia lub punktu skupienia wzroku w przestrzeni zadanego medium prezentującego bodźce wizualne. Stąd jakość danych zebranych podczas kalibracji rzutuje na jakość danych będących rezultatem procesu okulograficznego. Istotnym jest więc to, jak oprogramowanie kontrolujące procedurę kalibracji reaguje na pojawienie się potencjalnie niepoprawnych danych kalibracyjnych oraz jak ta procedura przebiega od strony współpracy z użytkownikiem.

Należy pamiętać, że sama postać modelu może być źródłem pogorszonej dokładności. Nadmierne dopasowanie lub zjawisko Rungego przy modelach zbyt wysokiego rzędu, prowadzą do pogorszenia skuteczności mapowania w częściach przestrzeni, w których nie występują punkty referencyjne. Z kolei wielomiany niższego rzędu mogą być niewystarczające, aby zapewnić odwzorowanie nieliniowości problemu. Szerzej problem doboru odpowiedniego modelu matematycznego opisano w Rozdziale 2.1.4.2.

### 2.2.1.3 Ruch, fizjologia i predyspozycje psychofizyczne

Niezależnie od typów i charakterów działania systemów okulograficznych, każdy z nich jest w jakimś stopniu czuły na zmianę położenia i orientacji głowy względem samej kamery rejestrującej obraz oczu albo docelowego medium w przestrzeni którego określane jest położenie punktu skupienia wzroku.

W przypadku wideookulografów opierających swoje działanie tylko na wykrywaniu położenia źrenicy (jak opisany w Rozdziale 2.1.5 *DIY*), należy zapewnić że



jedynym źródłem przemieszczenia środka źrenicy w obrazie z kamery jest ruch samych gałek ocznych względem głowy. Jakakolwiek zmiana względnego położenia kamery i głowy doprowadzi do błędnej interpretacji przesunięcia obrazu źrenicy i tym samym drastycznego zmniejszenia jakości danych, prowadzącego w efekcie do ich nieużyteczności. Skuteczne działanie jest możliwe dopiero, gdy dokona się kompensacji ruchów lub ponownie przeprowadzi się kalibrację.

Dopiero użycie stacjonarnego punktu referencyjnego związanego z głową, jakim w pewnym zakresie może być odbicie w powierzchni rogówki (por. Roz. 2.1.3.1), może pozwolić na pewną swobodę zmiany pozycji głowy [Duc07]. Przy czym należy pamiętać o dwóch związanych z tym problemach. Po pierwsze, o ile odnalezienie położenia źrenicy w obrazie z kamery jest obciążone błędem związanym z przetwarzaniem samej klatki pochodzącej z kamery (problem rozdzielczości, zaszumienie, rozmycie), o tyle dodatkowe uzależnienie systemu od odnalezienia położenia drugiej cechy obciążonej takimi samymi lub większymi problemami (odbicie rogówkowe zwykle jest znacznie mniejsze od źrenicy, por. Rys. 2.1), może prowadzić do jeszcze większego pogorszenia jakości danych źródłowych na których opiera się działanie mechanizmu estymacji kierunku spojrzenia. Drugą kwestią jest w takim przypadku wrażliwość na odległość pomiędzy kamerą, a okiem użytkownika. Oddalanie się powoduje skrócenie obrazu wektora pomiędzy środkiem źrenicy a odbiciem rogówkowym, co będzie błędnie interpretowane jako zmiana kierunku spojrzenia. Dopiero użycie obrazów obojga oczu pozwala na uwzględnienie kompensacji odległości na podstawie odległości źrenic. Przy czym tu należy też zwrócić uwagę na ruch konwergencyjny, który może znacząco zaburzyć tak prosty sposób szacowania odległości od kamery w przypadku, gdy są one niewielkie.

Problem jakości danych nie wynika wyłącznie z samego technicznego wymiaru środowiska eyetrackingowego. Źródłem niedoskonałości jest też element ludzki. Wynikające z ludzkich cech fizjologicznych czynniki prowadzące do pogorszenia jakości danych mogą mieć swoje źródło już na etapie kalibracji. W wyniku mikrodrgań gałki ocznej w trakcie trwania fiksacji (por. Roz. 1.2.2.2), nie jest możliwe trwałe skupienie się (skierowanie osi gałki ocznej) przez wymagany czas dokładnie na wyświetlanym markerze kalibracyjnym. Tym samym odwzorowanie mapujące pomiędzy położeniem istotnych cech w obrazie oka a punktami kalibracyjnymi, do wyliczenia którego służy kalibracja, nigdy nie będzie miało charakteru doskonałej precyzji. Zatem należy spodziewać się zaszumienia danych źródłowych już na etapie samej kalibracji, przez co warto rozpatrzyć pewnego rodzaju wstępne przetworzenie służące odrzuceniu skrajnie odstających próbek (ang. *outliers*), np. omawiane wcześniej obliczenie mediany położzeń [tob11].

Także mrugnięcia prowadzą do zaburzenia skuteczności wykrycia rzeczywistego położenia środka źrenicy. Stan, w którym powieki są częściowo przymknięte, może zostać zarejestrowany w trakcie zbierania osobnej klatki pochodzącej z kamery okulografu. Środek obszaru będącego widoczną częścią źrenicy może wówczas być przesunięty względem faktycznego jej środka, tym samym prowadząc do powstania próbki będącej przesuniętą względem pozostałych, zwiększając zaszumienie wśród próbek uczących. Dlatego ważne jest, by po ewentualnym wykryciu mrugnięcia lub wynikającej z jakiegokolwiek innego powodu utraty próbki, pominąć zebranie kilku następnych a także, jeśli to możliwe, kilku wcześniejszych. To daje także czas obser-

watorowi na ewentualne skorygowanie punktu skupienia wzroku, do którego mogło dojść podczas chwilowego zamknięcia oczu.

Zmęczenie wydaje się być kolejnym czynnikiem, który mógłby prowadzić do zwiększenia rozbieżności pomiędzy faktycznym punktem skupienia wzroku, a tym estymowanym przez system okulograficzny. Taka zależność miałaby wynikać z trudności w skupieniu się obserwatora podczas przechodzenia procedury kalibracyjnej. Co ciekawe, faktyczna zależność może być tutaj zupełnie przeciwna. Liczne badania wykazały bowiem wydłużający się czas fiksacji i wzmagające się opóźnienie sakad wraz z rosnącym zmęczeniem [LRSK13, APP<sup>+</sup>09, FPS<sup>+</sup>12], co paradoksalnie w przypadku charakteru kalibracji okulografu okazuje się pomocne. Tłumaczyć można to faktem, że zmęczony aparat wzrokowy jest mniej skory do wykonywania ruchów gałek ocznych, tym samym traktując spoglądanie na stacjonarne znaczniki jako swojego rodzaju odpoczynek.

Ostatecznie koniecznym jest wzięcie pod uwagę uwarunkowań konkretnych jednostek. Pewne predyspozycje i dysfunkcje psychofizyczne mogą bowiem znacząco utrudnić właściwą estymację kierunku spojrzenia czy też przejście samej procedury kalibracyjnej. Należy tutaj wyróżnić wszelkie odmiany oczopląsu i nystagmoidalnych ruchów gałek ocznych, powodujące znaczący rozrzut próbek kalibracyjnych. Jedyne w przypadku niewielkiej amplitudy i przy założeniu rozkładu normalnego drgań oczu, można oczekiwać pomyślnego działania systemu okulograficznego. Jaskra i inne przypadłości zaburzające czynności siatkówki poprzez wpłynięcie na zniekształcenie rozkładu jej czułości, mogą skutkować niezdolnością obserwatora do skupienia obrazu markera kalibracyjnego w sposób powtarzalny na powierzchni dna oka, prowadząc do nieskuteczności kalibracji. Także zaburzenia widzenia obucznego jak strabizm (zez) lub amblyopia mogą prowadzić do pogorszonej skuteczności działania okulografów binokularnych i, naturalnie, monokularnych, gdy te operują na oku które nie jest u danej osoby dominujące [Duc07].

## 2.2.2 Miara jakości systemu okulograficznego

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<sup>+</sup>11a].

### 2.2.2.1 Precyzja

Za [HNA<sup>+</sup>11b] precyzją określa się zdolność systemu eyetrackingowego do wskazywania tego samego kierunku, podczas ciągłej obserwacji tego samego punktu przez użytkownika.

Precyzja jest tym aspektem jakości, który zdaje się być w największej mierze wrażliwym na fizjologiczne cechy mechanizmów okomotorycznych. Paradoksalnie, hipotetyczny system okulograficzny, który cechowałby się idealną czułością na zmiany orientacji gałki ocznej, nie pada na wyjściu niezmiennego kierunku patrzenia. Błąd precyzji w takim wypadku i tak będzie zauważalny, będzie cechą indywidualną danego użytkownika i będzie zależeć od zakresu mikrodrgań gałki ocznej podczas

stabilnej fiksacji (por. Rozdział 1.2.2.2). Odfiltrowanie charakterystycznych dla problemów z precyzją, wysokoczęstotliwościowych szumów, z sygnału wyjściowego jest zatem zadaniem nie tylko kompensującym niedoskonałości samych systemów okulograficznych, ale też pozwalającym na poprawę użyteczności rezultatów obarczonych skutkiem fizjologicznego drżenia gałki ocznej.

Problem precyzji ma duże znaczenie także podczas procedury kalibracyjnej. Trzeba pamiętać, że zaszumienie tego rodzaju pojawia się również wtedy, gdy obserwator wpatruje się w punkty kalibracyjne. Prawidłowe obliczenie współczynników kalibracji będzie więc możliwe tylko wtedy, gdy uda się uzyskać rozkład normalny zebranych próbek, a więc gdy uda się odfiltrować błędne pomiary.

Sama miara precyzji nie jest wystarczająca do oceny jakości danych. Precyzja będzie wysoka nawet pomimo znacznego przesunięcia wskazywanego kierunku względem kierunku referencyjnego, jeśli tylko błędne wskazanie będzie stabilnie podtrzymywane.

### **2.2.2.2 Dokładność**

Dokładnością nazywa się średnią różnicę kątową pomiędzy referencyjnym kierunkiem patrzenia, a kierunkiem wskazanym przez eye tracker [HNA<sup>+</sup>11b]. Przy czym za referencyjny uznaje się fizyczny kierunek orientacji gałki ocznej, tj. wyznaczony przez odcinek łączący punkt obserwowanej sceny rzutowany na dołek środkowy oka ze środkiem obrotu oka.

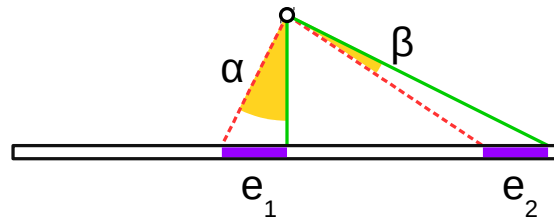
W połączeniu z miarą precyzji, dokładność obrazuje w intuicyjny sposób poprawność danych zebranych podczas pracy z systemem okulograficznym. Nie zapewnia ona jednak, że dane zebrane podczas następczej sesji przy tych samych warunkach będą miały podobną jakość.

### **2.2.2.3 Powtarzalność wyników**

Najbardziej rozmyty znaczeniowo z aspektów miary jakości systemu okulograficznego, dotyczy powtarzalności jakości uzyskiwanych rezultatów. Określenie to obejmuje zarówno zdolność systemu do zapewnienia zbliżonej jakości dla tego samego obserwatora przy kolejnych sesjach, jak i zdolność do zapewnienia porównywalnej jakości dla innych obserwatorów. To drugie ujęcie kwestii powtarzalności okazuje się być szczególnie istotne i problematyczne nawet dla najbardziej zaawansowanych rozwiązań eyetrackingowych. Zauważono dużą wariancję i dokładności, i precyzji pomiaru pomiędzy użytkownikami, którą próbuje się wiązać tak z różnicami w budowie fizjologicznej ich oczu, jak i z indywidualnymi predyspozycjami psychofizycznymi [tob11].

## **2.2.3 Pomiar jakości**

Aby pozyskać miarodajne wyniki stanowiące o jakości danych dostarczanych przez dany system okulograficzny, potrzebna jest ujednolicona, przemyślana procedura gromadzenia danych na potrzeby takiego pomiaru. Tylko wówczas można rozpatrywać porównywanie ze sobą skuteczności działania różnych systemów.



Rysunek 2.9: Zobrazowanie problemu malejącej miary kątowej błędu odpowiadającej błędowi na płaszczyźnie dla rosnących kątów obserwacji. Rysunek obrazuje dwa przypadki, w których faktyczny błąd na płaszczyźnie ekranu jest identyczny:  $e_1 = e_2$  (odległości oznaczone kolorem fioletowym). Natomiast różnica kątowa pomiędzy rzeczywistym kierunkiem obserwacji (ciągła, zielona linia), a kierunkiem wskazanym przez system okulograficzny (przerwana, czerwona linia) jest różna:  $\alpha > \beta$ . Czarny prostokąt wskazuje płaszczyznę ekranu, natomiast mały okrąg jest położeniem obserwatora.

Należy przy tym zwrócić uwagę, że dane okulograficzne, na podstawie których wyliczane będą miary dokładności i precyzji, winny mieć postać surową, niepoddaną jeszcze działaniu algorytmicznych metod poprawy jakości jak np. detekcja fiksacji (zob. Rozdział 2.2.4). Metody te mogą wszak być zastosowane opcjonalnie we wszystkich systemach bazujących na idei dostarczania kolejnych próbek w postaci współrzędnych kierunku widzenia bądź położenia punktu skupienia wzroku. Jeśli więc celem jest określenie i porównanie jakości danych dostarczanych przez dane środowisko, należy wykorzystać najwcześniejszą postać tychże danych, zależną wyłącznie od rozwiązań wbudowanych w samą procedurę wyznaczania kolejnych próbek.

Poniżej przedstawiono opis proponowanego przebiegu procedury pomiarowej, który został zastosowany w celu porównania jakości dwu rozwiązań okulograficznych: eye trackera *SensoMotoric Instruments RED250* oraz opisanego wcześniej *DIY* (por. Rozdział 2.1.5) [MB14].

### 2.2.3.1 Miara błędu

By umożliwić porównywanie pomierzonej jakości danych pomiędzy środowiskami, konieczne jest uniezależnienie uzyskanej miary od własności systemu wyświetlającego obraz podczas procedury badawczej. Niedopuszczalne byłoby wskazanie rezultatów w nawiązaniu do rozdzielczości ekranu lub rozmiaru bodźców bez uwzględnienia odległości użytkownika od płaszczyzny wyświetlania. Dlatego zakładając, że system okulograficzny dostarcza dane o położeniu punktu skupienia wzroku w postaci współrzędnych ekranowych, należy je przełożyć na miary kątowe związane z kątem widzenia nieruchomego obserwatora.

W dalszych rozważaniach kątowa miara błędu  $err$  wyznaczana będzie w poniższy sposób:

$$err = \arccos(\overline{v}_{gp} \circ \overline{v}_{tp}),$$

$$\overline{v}_{gp} = \left( \frac{(x_c - x_{gp}) \cdot w}{x_{res}}, \frac{(y_c - y_{gp}) \cdot h}{y_{res}}, d \right),$$

$$\overline{v}_{tp} = \left( \frac{(x_c - x_{tp}) \cdot w}{x_{res}}, \frac{(y_c - y_{tp}) \cdot h}{y_{res}}, d \right),$$

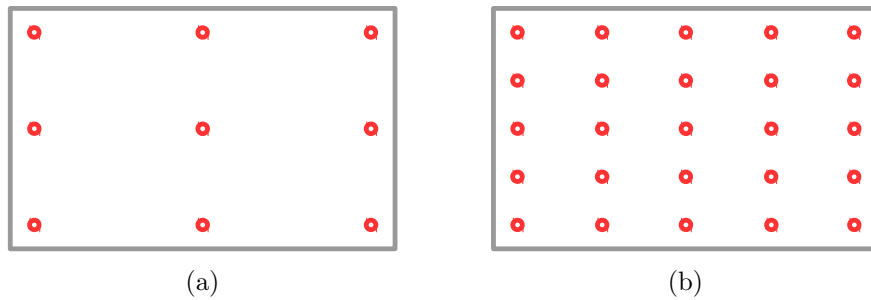
gdzie  $d$  jest odległością obserwatora od płaszczyzny ekranu o wymiarach obszaru roboczego  $w \times h$ , przy czym  $d$ ,  $w$  oraz  $h$  powinny być określone w tych samych, fizycznych jednostkach. Punkt  $(x_c, y_c)$  jest położeniem prostokątnego rzutu obserwatora na płaszczyznę ekranu,  $(x_{gp}, y_{gp})$  jest położeniem punktu skupienia wzroku otrzymanym z systemu okulograficznego, a  $(x_{tp}, y_{tp})$  jest położeniem punktu referencyjnego na który ma za zadanie spoglądać obserwator. Położenia dane są w jednostkach, w jakich wyskalowane jest wyjście z systemu okulograficznego, czyli najczęściej w pikselach. W tych samych jednostkach powinna też zostać określona pozioma ( $x_{res}$ ) oraz pionowa ( $y_{res}$ ) rozdzielczość roboczego obszaru ekranu. Przed obliczeniem iloczynu skalarnego ( $\circ$ ), wektory  $\overline{v}_{gp}$  oraz  $\overline{v}_{tp}$  winny zostać znormalizowane.

Należy zwrócić uwagę, że prosta kątowa miara błędu na płaszczyźnie maleje wraz ze zwiększaniem się odległości punktu referencyjnego od punktu, który jest rzutem położenia obserwatora na tę płaszczyznę (por. Rys. 2.9). Skuteczna miara kątowa powinna więc uwzględniać ten fakt i zapewnić, że rzut obserwatora na płaszczyznę pomiaru błędu będzie obrany jako środek obszaru.

### 2.2.3.2 Procedura pomiarowa

W przykładowej procedurze pomiarowej wykorzystano do porównania dwa systemy okulograficzne: zdalny, dostępny na komercyjnym rynku eye tracker *SensorMotoric Instruments RED250* oraz zbudowany niskim kosztem eye tracker *DIY* opisany w Rozdziale 2.1.5. Przy zbieraniu danych dla obu urządzeń, uczestnicy eksperymentu zajmowali tę samą pozycję w tej samej odległości (równej  $65\text{cm}$ ) od tego samego ekranu, co było kontrolowane użyciem podbródka. Starano się zachować kontrolowane warunki oświetleniowe, a kolejność wykonywania prób była losowo zmieniana, by możliwie zredukować wpływ czynników środowiskowych na pozyskane wyniki. W przypadku obu badanych sprzętów użyto procedur kalibracyjnych charakterystycznych dla nich, tj. opartych odpowiednio na oprogramowaniu *SMI iViewX* oraz *ITU Gaze Tracker*. Danymi na których oparto dalsze obliczenia były pobrane ze wspomnianego oprogramowania, wyliczone położenia punktów skupienia wzroku na płaszczyźnie ekranu, wyskalowane w pikselach.

Eksperyment rozpoczynał się od kalibracji w oparciu o 9-punktową siatkę położeń referencyjnych, pokrywającą równomiernie cały obszar roboczy 22-calowego monitora *Dell E2210* (por. Rys. 2.10a). Czas trwania procedury kalibracyjnej wynosił w obu przypadkach około 20 sekund. Na ekranie wyświetlony był punkt w postaci okrągłego markera, który przemieszczał się pomiędzy kolejnymi węzłami siatki, zatrzymując się w nich celem zebrania próbek. Uczestnik badania proszony był o uważne śledzenie spojrzeniem wyświetlanego punktu.



Rysunek 2.10: Rozłożenie punktów węzłowych na ekranie podczas kalibracji (a) oraz późniejszego zbierania danych służących ocenie ich jakości (b).

Analogicznie wyglądała właściwa część eksperymentu, związana ze zbieraniem danych służących później wyznaczeniu ich jakości. Tym razem jednak punktów referencyjnych rozrzuconych na płaszczyźnie ekranu było 25 (por. Rys. 2.10b), a procedura kontrolowana była w pełni przez autorskie oprogramowanie oparte o środowisko *MATLAB*. Czas trwania zbierania próbek dla pojedynczego punktu referencyjnego wynosił 2 sekundy. Próbkę zebraną przez pierwsze 800 milisekund były pomijane, aby upewnić się że obserwator będzie miał czas na skupienie się na nowym położeniu punktu. Dodatkowym ułatwieniem dla niego było to, że okrągły marker wskazujący punkt referencyjny przemieszczał się pomiędzy kolejnymi położeniami w sposób płynny, pozwalając na wygodne jego śledzenie. Również dane zebrane podczas ostatnich 200 milisekund ekspozycji danego punktu były pomijane w dalszych obliczeniach, by uniknąć wpływu ewentualnej predykcji ruchu na przemieszczenie punktu skupienia wzroku. Aby ułatwić skupienie się na wyświetlanym markerze, po zatrzymaniu w każdym z węzłów siatki jego promień malał z czasem.

Badanie oparto na danych zebranych od siedmiu obserwatorów w wieku 22 do 42 lat (średni wiek to 33 lata, w grupie były 3 kobiety i 4 mężczyźni). Każdy z uczestników badania powtórzył całą procedurę dziesięciokrotnie dla każdego z systemów okulograficznych. Zbieranie danych odbywało się na przestrzeni 15 dni. Badane osoby nie znały szczegółów procedury pomiarowej, choć były świadome celu prowadzonego eksperymentu.

Aby w pełni wyeliminować czynnik związany z fizjologią ludzkiego widzenia, ocena jakości działania systemu okulograficznego mogłaby zostać dokonana eksperymentalnie z użyciem urządzenia technicznego imitującego oko obserwatora, zdolnego do zmiany położenia podczas trwania pomiarów. Tak funkcjonujące *sztuczne oko* pozwoliłoby wykryć niedoskonałości działania samego systemu, pomijając wszelkie czynniki dotyczące drgań oczu i trudności ze skupieniem uwagi obserwatora [HNA<sup>+</sup>11a]. Z drugiej strony, tak uzyskane rezultaty odbiegałyby od rezultatów osiągniętych w praktycznych zastosowaniach i nie uwzględniałyby odporności systemu np. na mikroruchy gałek ocznych.

### 2.2.3.3 Rezultaty

Po uśrednieniu wyników pozyskanych w każdym z powtórzeń dla każdego z obserwatorów oraz dla każdego z 25 punktów referencyjnych, spodziewano się uży-

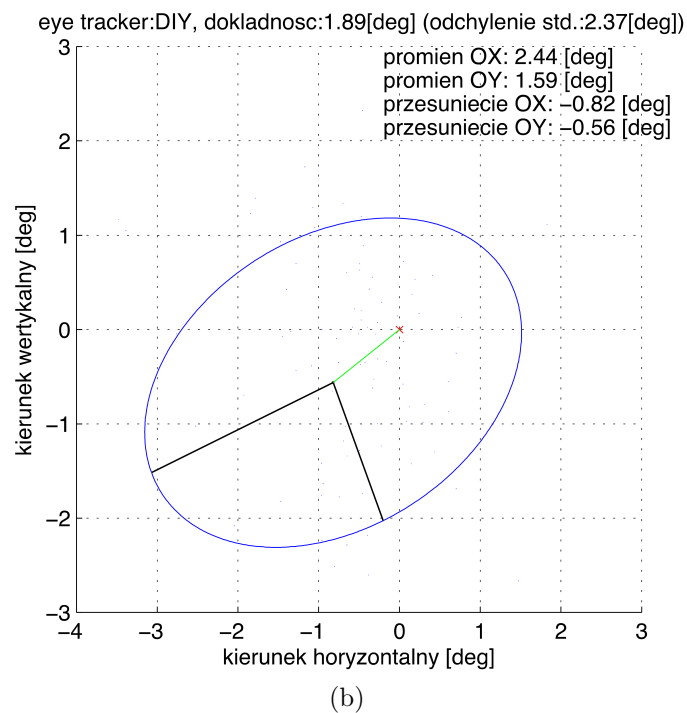
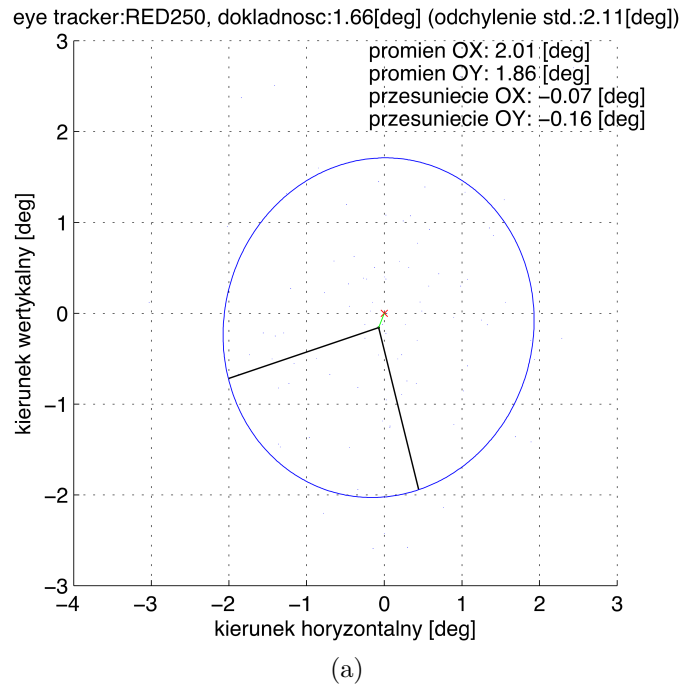
skać rezultat zbliżony do rozkładu normalnego. Otrzymany wynik zwizualizowano na Rys. 2.11 jako elipsy kowariancji rozkładów, po uprzednim wyliczeniu wartości w opisanej wcześniej mierze kątowej. Da się zauważyć, że rozkład odbiega od normalnego, co szczególnie wyraźnie widać w przypadku wyników dla eye trackera *DIY*. Elipsa w tym przypadku istotnie odbiega od kolistego kształtu, co sugeruje większą niedokładność w osi poziomej (poziomy promień elipsy wynosi  $2.44^\circ$ ), niż w osi pionowej (promień  $1.59^\circ$ ). W przypadku *SensoMotoric Instruments* wartości w poziomie i pionie są do siebie zbliżone (odpowiednio  $2.01^\circ$  i  $1.86^\circ$ ).

Warto zwrócić uwagę także na przesunięcie środka elipsy względem położenia punktu referencyjnego, które wskazuje na istnienie systematycznej tendencji w przypadku okulografu *DIY* (przesunięcie o  $-0.82^\circ$  w osi poziomej i  $-0.56^\circ$  w osi pionowej). Dla systemu *SensoMotoric Instruments RED250* takie przesunięcie jest znikome ( $-0.07^\circ$  w osi poziomej i  $-0.16^\circ$  w osi pionowej), co pozwala wnioskować o braku istotnej niejednorodności przesunięcia w obrębie całego ekranu. Należy jednak zwrócić uwagę na fakt, że badanie było prowadzone w warunkach unieruchomienia głowy badanej osoby. Można się spodziewać, że błędy dla eye trackera *SensoMotoric Instruments* będą większe gdy podczas korzystania z okulografu orientacja głowy będzie się zmieniała. Z geometrycznych aspektów systemu optycznego na bazie którego działa urządzenie można też wnioskować, że przy obserwacji obrazu pod dużym kątem w stosunku do kierunku prostopadłego do ekranu, może pojawić się znaczące, systematyczne przesunięcie otrzymywanych próbek.

Przyczyny systematycznego przesunięcia wyników otrzymywanych z eye trackera *DIY* można doszukiwać się w jego wrażliwości na ruchy użytkownika. Pomimo wykorzystania podbródka stabilizującego położenie głowy, nawet niewielkie jej ruchy powstające w wyniku oddychania lub zmęczenia mogą wpłynąć na powstanie takiego zjawiska w kolejnych próbkach. Dodatkowo mogło nastąpić powolne, grawitacyjne, nieznaczące zsuwanie się ramy okularów eye trackingowych w dół nosa użytkownika, co też tłumaczyłoby przesunięcie w osi pionowej.

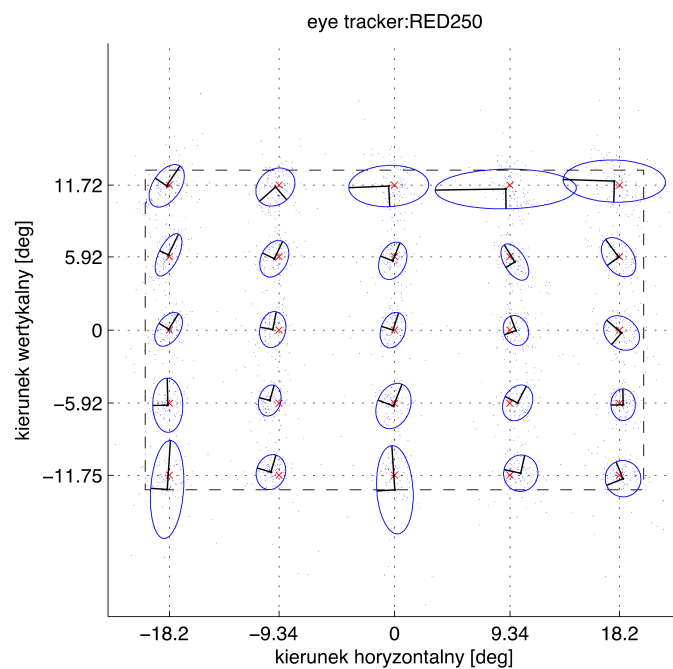
Przyglądając się uśrednionej kowariancji rozkładów z uwzględnieniem podziału na punkty referencyjne (por. Rys. 2.12), można zaobserwować większą niedokładność zachodzącą dla punktów położonych przy krawędziach ekranu. Szczególnie dobrze jest to widoczne dla górnych punktów ( $+11.72^\circ$ ) w przypadku eye trackera *SensoMotoric Instruments*, który działa zdalnie. Tym samym musi opierać się na obrazach zarówno źrenicy, jak i odbić rogówkowych, przy czym te drugie mają znacznie mniejszy rozmiar w klatkach rejestrowanych z użyciem kamery. Biorąc pod uwagę także umieszczenie urządzenia poniżej ekranu, można tłumaczyć to zjawisko utrudnionym rozpoznawaniem położenia istotnych cech obrazu oczu w wyniku ich przysłonięcia powiekami przy dużym kącie obserwacji.

Kolejnym aspektem jakości danych jest jej powtarzalność w uzależnieniu od obserwatora. Na tym polu zebrane dane pozwoliły pokazać, że jakość jest w bardzo znaczącym stopniu powiązana z konkretnym użytkownikiem systemu. Przy czym w kolejnych powtórzeniach dla danej osoby uzyskuje się zbliżone rezultaty. Na Rys. 2.13 przedstawiono średnie wartości błędów i przedziały ufności dla poszczególnych użytkowników biorących udział w badaniu. Średnia dokładność *SensoMotoric Instruments* waha się od  $1.18^\circ$  do  $2.89^\circ$ , a *DIY* od  $1.07$  do  $3.5$  stopnia, zależnie od osoby.

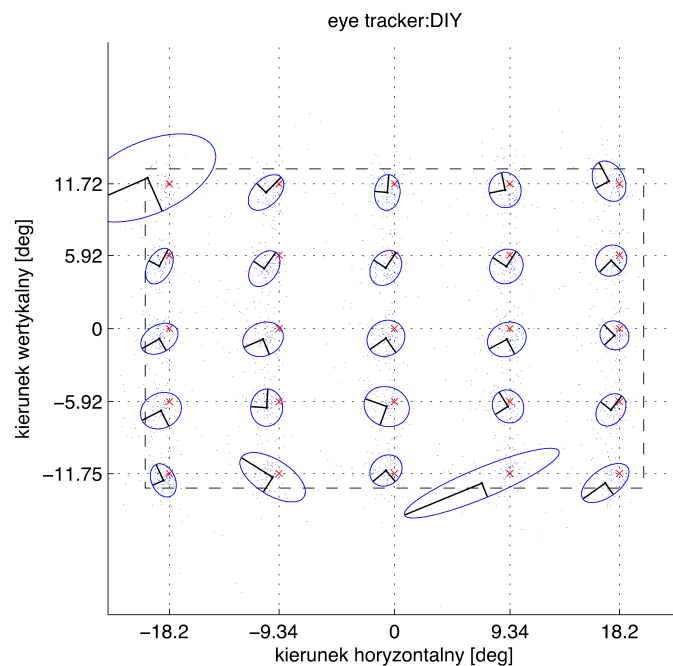


Rysunek 2.11: Wizualizacja kowariancji usrednionego błędu położenia próbek dla wszystkich obserwatorów, wszystkich powtórzeń i wszystkich punktów referencyjnych odpowiednio dla *SensoMotoric Instruments RED250* (a) i *DIY* (b). Czerwonym krzyżykiem oznaczono położenie usrednionego punktu referencyjnego.



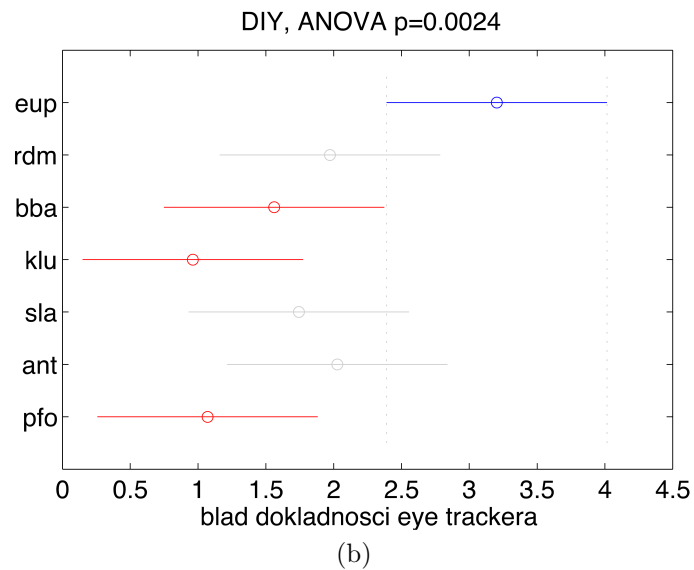
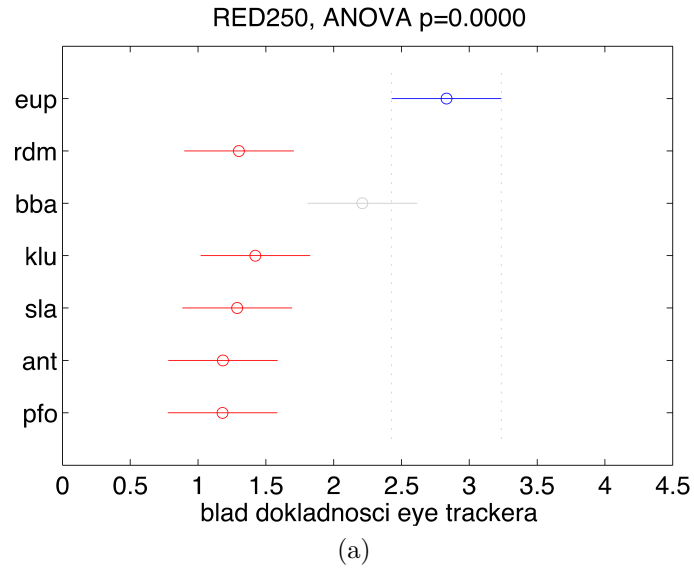


(a)



(b)

Rysunek 2.12: Wizualizacja kowariancji uśrednionego błędu położenia próbek dla wszystkich obserwatorów, wszystkich powtórzeń z uwzględnieniem rozkładu punktów referencyjnych odpowiednio dla *SensoMotoric Instruments RED250* (a) i *DIY* (b). Czerwonym krzyżykiem oznaczono położenie punktów referencyjnych.



Rysunek 2.13: Analiza wariancji średniego błędu dokładności badanych okulografów dla poszczególnych obserwatorów. Średnia wartość błędu zaznaczona jest okręgiem. Poziome odcinki pokazują wielkość przedziałów ufności. Wyniki odpowiednio dla *SensoMotoric Instruments RED250* (a) i *DIY* (b). Kolorem niebieskim oznaczono tych obserwatorów, dla których wielkość błędu znacząco przewyższała średnią, natomiast kolorem czerwonym tych obserwatorów, dla których błąd był niższy.

#### 2.2.3.4 Wnioski

Dla eye trackera *DIY* szczególnie ważne jest wyeliminowanie wrażliwości na ruchy głowy użytkownika. Wiąże się to z zadbaniami, by badana osoba faktycznie pozostawała bez ruchu podczas procedury kalibracyjnej i całego badania. Innym, słusznym podejściem do rozwiązania tej kwestii byłoby też wprowadzenie mechanizmu korygującego błąd systematyczny wynikający ze zmiany orientacji i położenia głowy, bazującego na zasadach śledzenia głowy (ang. *head tracking*, zob. Rozdz. 2.1.5).

O ile urządzenie *SensoMotoric Instruments* zdaje się dostarczać dane obciążone mniejszym błędem w stosunku do *DIY*, co wynika z braku wyraźnego, systematycznego przesunięcia, to warto zwrócić uwagę na porównywalną precyzję obu urządzeń która objawia się zbliżonymi promieniami elips kowariancji. Dla *DIY* nie występuje też tak wyraźne pogorszenie jakości przy krawędziach ekranu (poza skrajnymi wychyleniami w niektórych narożnikach), co wynika z małej odległości kamery od oka, a tym samym wyraźnego, dużego obrazu źrenicy dla większego zakresu kątów widzenia.

Należy też zwrócić uwagę, że pomierzone wartości błędów znacząco różnią się od deklarowanych przez producenta urządzenia *SensoMotoric Instruments*. Pokazuje to jak bardzo ważne jest wprowadzenie i upowszechnienie ujednoczonych, sprawnych i dokładnych metod pomiaru jakości danych dostarczanych przez systemy okulograficzne.

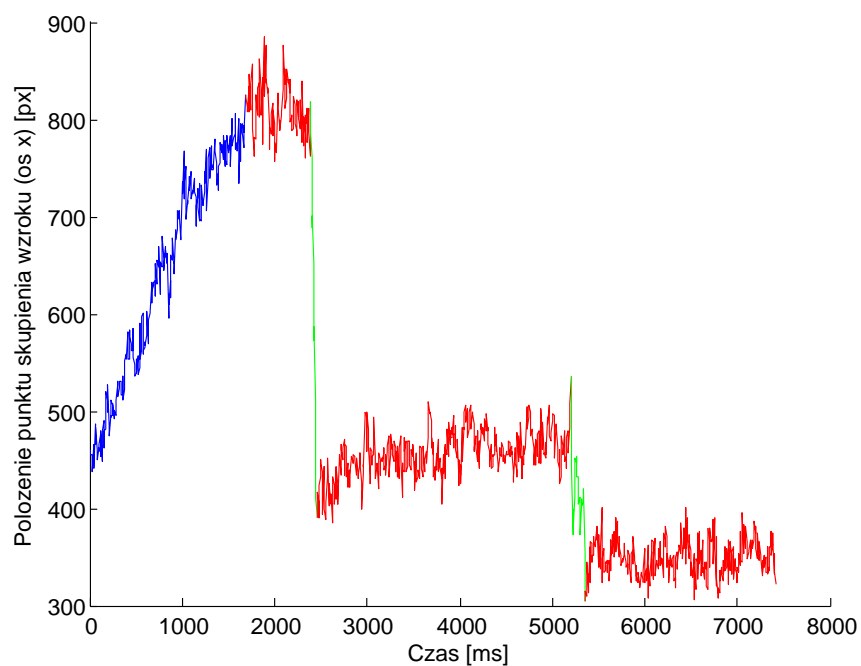
Badanie pokazało, że konieczne jest poddawanie danych otrzymanych z okulo grafu odpowiedniej filtracji i przekształceniu, by podjąć próbę kompensacji niedoskonałości wynikających z opisanych wyżej przyczyn błędów. Zanim dane zostaną skonsumowane przez końcową aplikację lub zostaną wykorzystane w ostatecznym badaniu, powinny zostać przetworzone z użyciem technik które poprawią ich jakość. Takie techniki będziemy zbiorczo nazywać technikami filtracji danych okulograficznych.

#### 2.2.4 Filtracja danych

Zaszumienie sygnału, będące cechą charakterystyczną danych obciążonych błędem precyzji, jak i systematyczne jego przesunięcie względem rzeczywistego kierunku spojrzenia związane z małą dokładnością, są zjawiskiem powszechnym w okulo grafii. Przykładowy sygnał otrzymany z urządzenia do śledzenia wzroku przedstawiono na Rysunku 2.14.

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 za-



Rysunek 2.14: 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.

stosowaniach okulograficznych bierze się pod uwagę przede wszystkim fiksacje oraz sakady, pierwsze z nich przyjmując za czas świadomego percypowania obrazu, a drugie odfiltrując 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 opisane niżej algorytmy *I-DT* oraz *I-VT* [SG00].

#### 2.2.4.1 Metody *I-DT* i *I-VT*

Obie te metody mają na celu wyodrębnienie punktów fiksacji z dwuwymiarowego sygnału pochodzącego z okulografu. Idea obu opiera się jednak na odmiennych definicjach fiksacji.

Metoda *I-DT* (ang. *Identification - Dispersion Threshold*) uznaje za fiksacje takie zbiory kolejnych próbek, które przez przynajmniej zadany czas  $t$  mają dyspersję o co najwyżej wielkości  $d$ . Przy czym dyspersja jest rozumiana jako rozrzut położeń punktów skupienia wzroku na płaszczyźnie lub rozrzut kątów kierunku widzenia, najczęściej mierzonych jako odległość od centroidu. Zakres dyspersji często nazywany jest jej *oknem*. Literatura wzmiankuje różne wartości progowe  $d$  i  $t$ , dotyczą one jednak zazwyczaj dyspersji kątowej na poziomie  $1^\circ - 3^\circ$  i czasów rzędu 100 – 400 milisekund [BRG81, SE80, WK81]. Dobór konkretnych wartości powinien zależeć zarówno od cech danego systemu okulograficznego jak jego częstotliwość pracy, opóźnienie, dokładność, jak i od charakteru zadania postawionego obserwatorowi [Wid84, Bli09].

Implementacje tej metody mogą różnić się sposobem obliczania wyjściowej wartości. Można spotkać implementacje czasu rzeczywistego które za wyjściowe położenie punktu fiksacji przyjmują pierwszą próbkę nowo wykrytej fiksacji, centroid próbek z pierwszego okna pozwalającego na wykrycie fiksacji [Sen09], albo też dynamicznie aktualizują położenie wraz z otrzymywanymi nowymi próbkami które wciąż przynależą do bieżącego okna dyspersji – również w dwojaki sposób: albo uwzględniając całość próbek danej fiksacji, albo tylko poruszające się w czasie okno. W implementacjach *offline* natomiast zazwyczaj spotyka się statyczne, nieruchome końcowe położenie, wyliczane na podstawie wszystkich próbek zakwalifikowanych do danej fiksacji.

Ciekawym rozwinięciem metody opartej o dyspersję jest algorytm, który zakłada wykorzystanie minimalnych drzew rozpinających *I-MST* (ang. *Identification - Minimum Spanning Trees*). Myślą przewodnią tej metody jest takie utworzenie drzewa złożonego ze zbiorów próbek, które zapewni najmniejszą całkowitą liczbę połączeń [SG00]. Nie jest to jednak popularnie wykorzystywane podejście.

Metoda *I-VT* (ang. *Identification - Velocity Threshold*) opera się na uznawaniu za sakady takich fragmentów sygnału, gdzie kątowa szybkość zmiany kierunku widzenia przekracza wartość progową  $v$  przez przynajmniej przez czas  $t$ . Kryterium czasowe jest tutaj istotne, aby za sakady nie były uznawane piki związane z zaszumieniem danych. Wartość progowa  $v$  określana jest w literaturze na poziomie  $15^\circ/s$ ,

natomiast minimalny czas trwania sakady na 15 milisekund [EV95]. Podobnie jak w przypadku metody *I-DT*, wartości powinny jednak zostać dobrane do konkretnego środowiska okulograficznego i charakteru zadania, podczas którego zbierane będą informacje o kierunku patrzenia [Kar00]. Przy niewielkiej częstotliwości próbkowania systemu okulograficznego, określenie czasu trwania sakady może być niemożliwe (dla 30 Hz czas trwania sakady na poziomie 15 ms będzie dotyczył pojedynczej klatki).

Probabilistyczna klasyfikacja fiksacji i sakad może się odbywać również z użyciem dwustanowych ukrytych modeli Markowa, których użycie stanowi rozszerzenie podstawowych założeń opisanej wyżej metody *I-VT*. Idea takiego podejścia opiera się na różnym rozkładzie prędkości dla tych dwóch rodzajów ruchów oczu. Mówi się wówczas o metodzie *I-HMM*. Choć dokładność danych ulega wówczas poprawie, problemem są szybkości zbliżone do wartości progowej, które znacząco utrudniają poprawne rozpoznawanie fiksacji i sakad [SG00].

#### 2.2.4.2 Problem identyfikacji obiektu uwagi

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].

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 płynnych pościgów (ang. *smooth pursuits*), wykonywanych przez 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<sup>+</sup>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.

## 2.3 Metoda GDOT

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 dostarczanych 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].

### 2.3.1 Algorytm identyfikacji

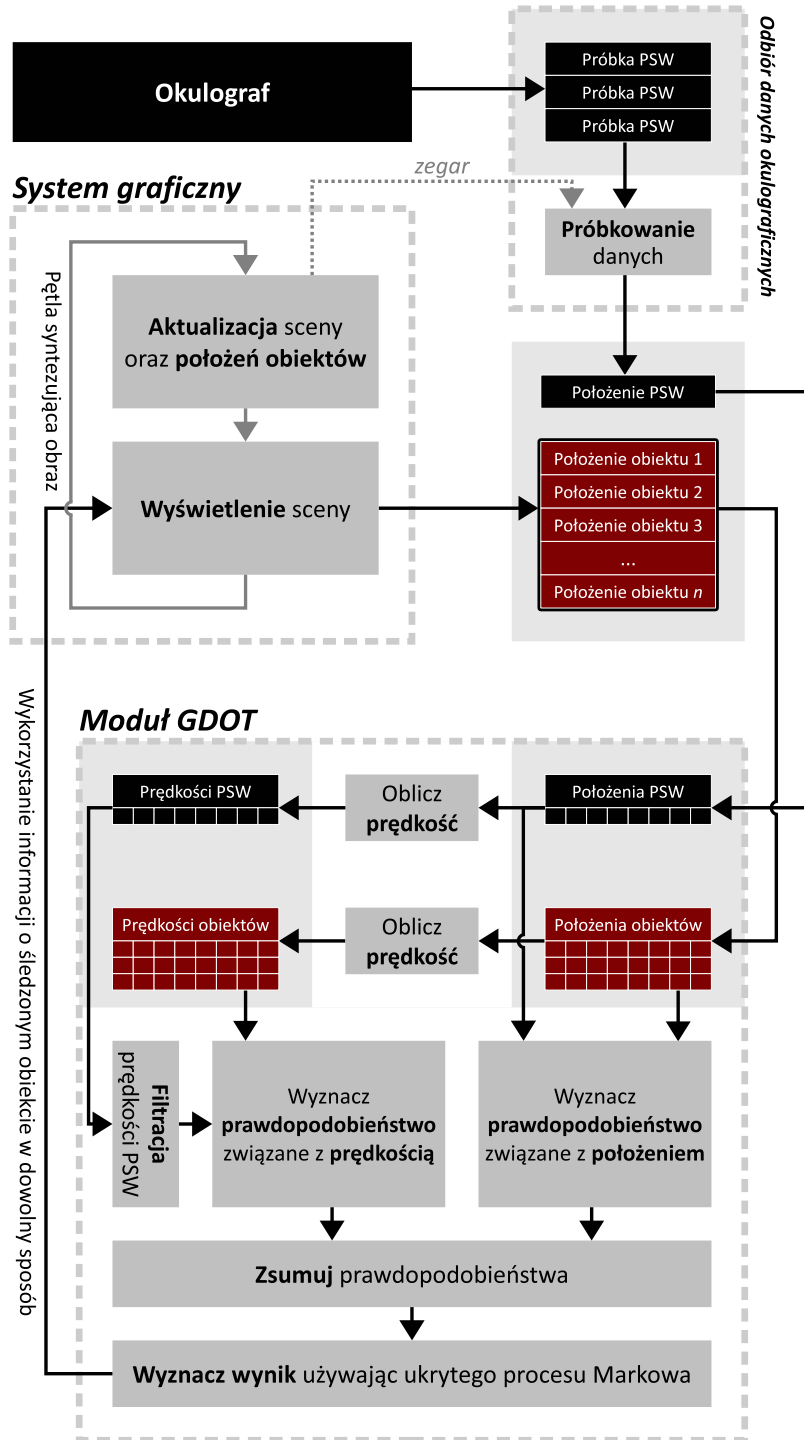
#### 2.3.1.1 Podstawowe założenia

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 (zob. Rozdz. 1.2.2.3). 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.

Tym samym ważną przesłanką pozwalającą na rozróżnienie sytuacji, w których śledzone są dwa poruszające się obiekty, staje się podobieństwo ich prędkości do prędkości przemieszczania się punktu skupienia wzroku. Aby możliwe było ich porównanie, prędkości te powinny zostać obliczone w tej samej przestrzeni. Z racji tego, że wspólną cechą dla systemów komputerowych syntezy obraz w czasie rzeczywistym jest uzyskiwanie na jakimś etapie położenia wyświetlanych obiektów w kartezyjskiej, dwuwymiarowej przestrzeni ekranu, jak i większość systemów okulograficznych dostarcza dane w tejże, zdecydowano się obrać tę właśnie przestrzeń jako służącą rozwiązywaniu problemu postawionego przed omawianym algorytmem. Nic nie stoi jednak na przeszkodzie, by ideę metody przenieść do jakiegokolwiek innej przestrzeni pozwalającej na jednoznaczne wyznaczenie położenia oraz prędkości punktu. Praktycznym przykładem może być tutaj operowanie układem współrzędnych sferycznych, bazującym na kątach widzenia człowieka w przestrzeni trójwymiarowej.

Rysunek 2.15: Schematyczne przedstawienie sposobu działania algorytmu identyfikacji obiektu uwagi *GDOT*. Skrótom *PSW* oznaczono Punkt Skupienia Wzroku.





Rysunek 2.16: Wizualizacja przykładowego rozmieszczenia potencjalnych odbiorców uwagi na trójwymiarowej scenie.



### 2.3.1.2 Zbiór potencjalnych odbiorców uwagi

*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.

Jako że podejście oparte o skończony zbiór obiektów traktuje każdy element sceny jako rodzaj singletonu, pojedynczego bytu określonego jednym położeniem, w przypadku dużych elementów niezbędne może okazać się rozmieszczenie w ich obrębie więcej niż jednego takiego obiektu. Przykład takiej sytuacji pokazano na Rys. 2.16, gdzie każda ze ścian pokryta jest siatką takich punktowych obiektów, a figura smoka składa się z kilku związanych z semantycznie istotnymi jego częściami (głowa, ogon, tułów).

Warto zwrócić uwagę, że w dynamicznej scenie, gdzie kamera i obiekty pozostają w ruchu, może dochodzić do czasowego przysłonięcia niektórych elementów, a co za tym idzie potencjalnych odbiorców uwagi. Człowiek jest jednak w stanie śledzić z pomocą płynnego pościgu także te obiekty, które przez pewien czas, znikając za przeszkodami, pozostają niewidoczne w jego polu widzenia [BF85]. Dokonywana jest wówczas predykcja trajektorii ruchu. Aby system dopuszczał wykrywanie także takich etapów śledzenia i nie gubił obiektów, gdy te zostaną przysłonięte, ze zbioru potencjalnych odbiorców nie powinno się usuwać tych elementów, co do których istnieje podejrzenie, że mogą być dalej obserwowane i mogą powrócić w obszar widzenia obserwatora.

Podstawowa implementacja algorytmu *GDOT* zakłada, że w każdej chwili je-

go działania aktywny, a więc obserwowany przez użytkownika, jest dokładnie jeden element ze zbioru potencjalnych odbiorców uwagi. Nie jest więc przewidziany stan, w którym nie występuje ten najbardziej prawdopodobny obiekt, na którym miałyby spoczywać wzrok obserwatora. Podejście takie może być w niektórych zastosowaniach niepożądane, toteż wprowadzenie stanu wolnego zdaje się być interesującym, aczkolwiek nieobjętym niniejszą rozprawą elementem dalszych prac nad algorytmem.

### 2.3.1.3 Odbiór danych z okulografu

Dane dostarczane przez sprzęt okulograficzny mają postać próbek zawierających współrzędne chwilowego estymowanego położenia punktu skupienia wzroku w przestrzeni, w której wcześniej przeprowadzona została kalibracja. Oprócz tego wiele *eye trackerów* oferuje szeroką gamę innych informacji, które mogą zostać dołączone do każdej z próbek. Wymienić tutaj należy szczególnie znacznik czasowy, pośrednie współrzędne punktów charakterystycznych na obrazie oka (środky źrenicy, odbicia rogówkowe), aktualną odległość pomiędzy źrenicami czy też binarną informację o odnalezieniu obrazu każdego z oczu w klatce otrzymanej z kamery [Sen09, Sen10, Tob09]. Z punktu widzenia metody *GDOT* istotne są jednak tylko współrzędne ekranowe punktu skupienia wzroku w każdej z pochodzących z okulografu próbek. Metoda może więc być stosowana z praktycznie każdym współczesnym rozwiązaniem okulograficznym.

Kluczową kwestią związaną z odbiorem danych jest zagadnienie synchronizacji pomiędzy częstotliwością pracy urządzenia do śledzenia wzroku, a częstotliwością pracy systemu korzystającego z wyjścia metody *GDOT* oraz zasilającego go danymi o prezentowanym użytkownikowi obrazie. Należy rozpatrzyć tutaj dwie możliwe niezgodności: gdy system prezentujący obraz i konsumujący wynik działa wolniej, bądź gdy działa on szybciej od okulografu.

Częstotliwość dostarczania próbek przez *eye tracker* najczęściej jest związana z liczbą klatek na sekundę, które są rejestrowane przez jego kamerę. Należy się więc w tej kwestii spodziewać, zależnie od sprzętu, bardzo dużych częstotliwości nawet rzędu kilkuset  $Hz$  [Sen09], albo niewielkich na poziomie  $30Hz$  jak w systemach niskokosztowych [MKNB12]. Z kolei właściwa praca algorytmu *GDOT* wiązana jest zazwyczaj w sposób synchroniczny z cyklem działania systemu konsumującego jego wyjście. Tym samym można się spodziewać, że dla systemów grafiki czasu rzeczywistego będzie to wartość zbliżona do  $60Hz$  w przypadku syntezy obrazu monoskopowego lub  $120Hz$  w przypadku obrazu stereoskopowego.

Istotną częścią systemu wykorzystującego *GDOT* jest więc moduł pozwalający na buforowanie i interpolowanie dyskretnego sygnału pochodzącego z okulografu, który umożliwi efektywne i skuteczne wprowadzanie do algorytmu danych wejściowych pochodzących z dwu nieskoordynowanych źródeł. Zazwyczaj wystarczające jest przeprowadzenie interpolacji liniowej współrzędnych w przestrzeni ekranu, co zostało poczynione na potrzeby wszystkich opisanych w niniejszej pracy badań.

### 2.3.1.4 Przetworzenie danych wejściowych

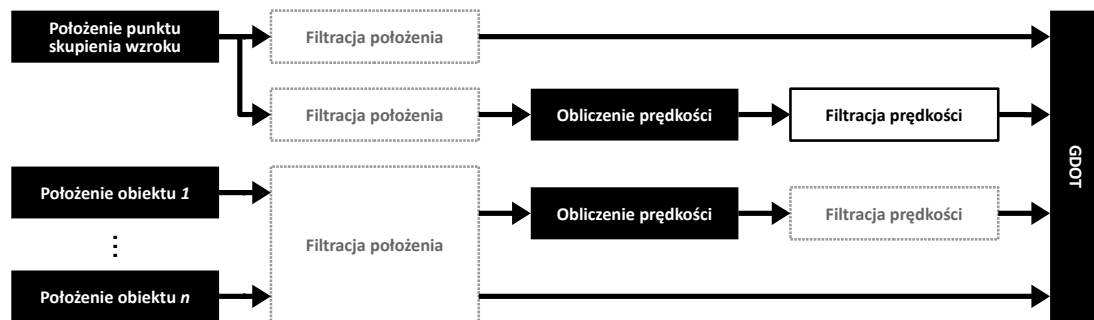
O ile opisany wyżej moduł synchronizacji częstotliwości próbkowania danych okulograficznych z częstotliwością pracy algorytmu umiejscowiony jest w formal-

nej architekturze proponowanego rozwiązania poza główną częścią *GDOT* (por. Rys. 2.15), o tyle dane wejściowe przekazane do tej części wymagają też odpowiedniej obróbki już po ich wprowadzeniu. Na tym etapie przyjmujemy już, że każdej jednej próbce pochodzącej z okulografu odpowiada dokładnie jedna próbka zawierająca aktualne położenia obiektów ze zbioru odbiorców uwagi.

Wyjście algorytmu zależy nie tylko od chwilowego stanu systemu, lecz od skończonej części jego historii, co dalej będzie nazywane oknem jego działania. Z tego powodu konieczne jest przechowanie historii zarówno danych o współrzędnych punktu skupienia wzroku, jak i danych o położeniach potencjalnych odbiorców uwagi w takiej ilości, aby pokryły czas trwania (*szerokość*) okna. Bufor realizujący to postanowienie ma dokładnie taką samą pojemność próbek dla danych okulograficznych, jak i danych o scenie.

Na potrzeby algorytmu *GDOT* konieczne jest obliczenie prędkości chwilowych dla każdej z próbek, aby możliwe było ich porównanie ze sobą w obrębie każdego kroku czasowego. Należy wspomnieć, że dla poprawienia efektywności działania algorytmu podczas pracy w trybie *on-line*, warto wyniki tych obliczeń również buforować, jako że prędkości z ubiegłych kroków czasowych nie będą ulegały zmianie wraz z podawaniem na wejście nowych próbek.

Zarówno położenia kolejnych punktów skupienia wzroku, jak i prędkości wyliczone na ich podstawie, mogą być silnie zaszumione. Zależnie od cech danego środowiska, należy rozpatrzyć filtrację danych w celu usunięcia zakłóceń mogących zniekształcić uzyskiwanie poprawnych wyników. Na Rys. 2.17 pokazano miejsca, w których możliwe jest zastosowanie filtracji dolnoprzepustowej. W omawianej dalej implementacji posłużono się jedynie prostym filtrem pudełkowym (ang. *box filter*) dla prędkości punktu skupienia wzroku. Położenia i prędkości obiektów sceny uznano za niewymagające filtracji, gdyż pochodzą ze stabilnego systemu syntezującego obraz.



Rysunek 2.17: Miejsca w przepływie danych wejściowych algorytmu *GDOT*, gdzie można zastosować filtrację w celu eliminacji szumów. Linia ciągłą oznaczono etap, na który został użyty filtr dolnoprzepustowy w implementacji opisywanej w niniejszej rozprawie.

### 2.3.1.5 Obliczenie prawdopodobieństw

Aby można było uznać, że dany obiekt jest śledzony przez obserwatora, estymowany przez system okulograficzny punkt skupienia wzroku musi znajdować się odpowiednio blisko tego obiektu, a prędkość ich przemieszczania się powinna być zbliżona. Na podstawie tego założenia można opisać prawdopodobieństwo śledzenia obiektu  $o_i$  następującą formułą:

$$P(o_i) = P(p_i \cup v_i) = 1 - (1 - P(p_i))(1 - P(v_i)).$$

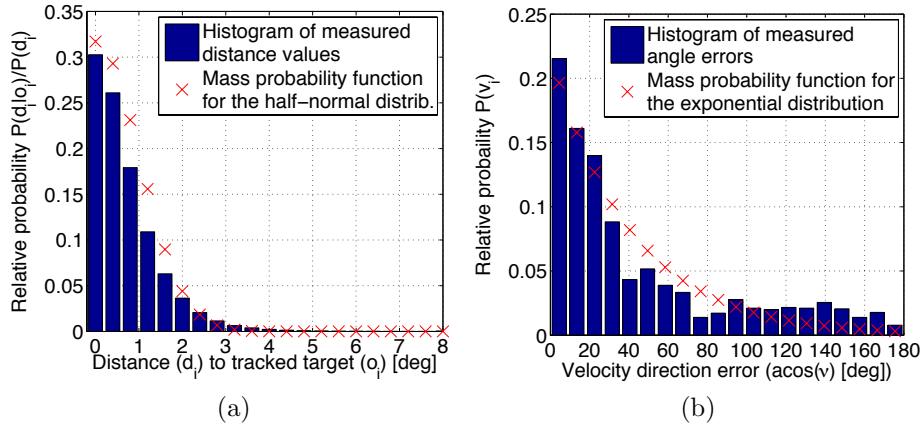
W powyższym zapisie  $P(p_i)$  oznacza prawdopodobieństwo, że dana próbka z okulografu dotyczy sytuacji, w której położenie punktu skupienia wzroku jest związane z obiektem  $o_i$ .  $P(v_i)$  natomiast jest prawdopodobieństwem tego, że ścieżka przemieszczania się punktu skupienia wzroku pokrywa się z torem ruchu obiektu, co wyznaczone jest przez podobieństwo ich chwilowych wektorów prędkości.

Należy spodziewać się, że częstym zjawiskiem jest niezgodność przynajmniej jednego z dwu elementów, położenia lub prędkości punktu skupienia wzroku, z rozpatrywanym obiektem. W przypadku położenia, przyczyną niezgodności są najczęściej zaszumienie sygnału oraz niedokładność kalibracji systemu. Ta niedokładność nie musi być jednorodna w obrębie całej przestrzeni ekranu i w czasie. Z kolei w przypadku prędkości, niezgodność jest największa dla obiektów stacjonarnych, gdzie pomierzone prędkości próbek z okulografu wiążą się z zaszumieniem sygnału, a nie faktycznym przemieszczaniem się punktu skupienia wzroku. Jednak gdy tylko obserwowany obiekt zaczyna się poruszać, da się wskazać bardzo wysoką zgodność w oparciu o prędkości związaną z inicjacją płynnego pościgu wykonywanego przez gałki oczne. Z powodu wskazanych problemów, których przyczyny opisano szerzej w Rozdziale 2.2, końcowe prawdopodobieństwo jest obliczane jako suma cząstkowych prawdopodobieństw, a nie ich iloczyn. W ten sposób możliwe jest wykorzystanie zalet obu cząstkowych prawdopodobieństw w korzystnych dla nich warunkach, jednocześnie nie wykluczając przypadku, gdy jedno z nich jest niskie.

Aby matematycznie zamodelować związane z położeniem prawdopodobieństwo  $P(p_i)$  trafienia w obiekt, posłużono się funkcją probabilistyczną opartą o założenie, że rozkład błędów danych okulograficznych dotyczący położenia punktu skupienia wzroku ma kształt połowy rozkładu normalnego (por. Rys. 2.18a):

$$P(p_i) = \frac{P(d_i|o_i)P(o_i)}{P(d_i)} = \omega_p \exp\left(\frac{-d_i^2}{2\sigma_s^2}\right).$$

Za pomocą  $d_i$  oznaczono odległość pomiędzy położeniem estymowanego punktu skupienia wzroku, a położeniem rozpatrywanego obiektu. Dla przypadku w którym identyfikacji obiektu uwagi dokonuje się w obrębie przestrzeni ekranu, odległość ta może być odległością euklidesową między dwoma punktami na płaszczyźnie.  $\sigma_s$  jest parametrem związanym z błędem położenia danego środowiska okulograficznego, odpowiada więc za kształt wykresu rozkładu normalnego przedstawionego na Rys. 2.18a. Wartość tego parametru powinna zostać dobrana dla konkretnych warunków technicznych, w których wykorzystywany jest algorytm. Z kolei  $\omega_p$  jest parametrem określającym udział tego cząstkowego prawdopodobieństwa w obliczaniu



Rysunek 2.18: Rozkłady błędów dla położenia punktu skupienia wzroku (a) i prędkości punktu skupienia wzroku (b) otrzymywanych z systemu okulograficznego. Na podstawie danych zebranych podczas badań związanych z wyznaczeniem metryki oceny jakości (zob. Rozdział 2.3.2.1).

końcowego prawdopodobieństwa  $P(o_i)$  i jego wartość powinna zawierać się w przedziale  $< 0, 1 >$ .

Głównym celem zastosowania cząstkowego prawdopodobieństwa związanego z prędkością jest uwzględnienie śledzenia obiektów z użyciem płynnego pościgu. Ten ruch gałek ocznych nie jest procesem chwilowym i odbywa się najczęściej w okresach dłuższych niż 100 milisekund (por. Rozdz. 1.2.2.3), co umożliwia wygładzenie sygnału obliczonej prędkości w celu eliminacji nadmiernego zaszumienia. Z tego właśnie powodu w rozpatrywanej implementacji algorytmu *GDOT* zastosowano filtr prędkości wskazany na Rys. 2.17.

Dla wyznaczenia prawdopodobieństwa  $P(v_i)$  zbieżnej prędkości wykorzystano założenie o rozkładzie błędu zbliżonym do rozkładu wykładniczego (por. Rys. 2.18b):

$$P(v_i) = \frac{P(v_i|o_i)P(o_i)}{P(v_i)} = \omega_v \exp\left(\frac{-\arccos(v)}{\sigma_v}\right).$$

Podobnie jak w przypadku prawdopodobieństwa powiązanego z położeniem, tutaj też wprowadzono parametr  $\sigma_v$  dotyczący charakterystycznej wielkości błędu dla danego środowiska okulograficznego. Również analogicznie występuje tutaj  $\omega_v$  będące współczynnikiem udziału tego prawdopodobieństwa cząstkowego przy obliczaniu prawdopodobieństwa końcowego.

Z uwagi na zaszumienie sygnału na potrzeby działania algorytmu pominięto długości wektorów prędkości punktu skupienia wzroku i rozpatrywanego obiektu, biorąc pod uwagę wyłącznie ich względną orientację, tj. kąt jaki tworzą na płaszczyźnie. Zatem  $v$  w powyższym wzorze określone jest następująco:

$$v = \frac{u_t \cdot v_{t,i} + \epsilon}{\|u_t\| \|v_{t,i}\| + \epsilon},$$

gdzie  $u_t$  jest chwilowym, poddanym wygładzeniu wektorem prędkości punktu skupie-

nia wzroku, a  $v_{t,i}$  jest chwilowym wektorem prędkości obiektu  $o_i$ . Stała  $\epsilon$  jest małą wartością zapobiegającą dzieleniu przez zero w przypadku gdy obiekt lub punkt skupienia wzroku miałyby pozostać w bezruchu.

### 2.3.1.6 Wyjście algorytmu

Na wyjściu implementacji algorytmu *GDOT* pojawia się dyskretna wartość będąca identyfikatorem obiektu, który został uznany za najbardziej prawdopodobnego odbiorcę uwagi wizualnej w danej chwili. Przy prostej implementacji proponowanego algorytmu, wartość ta może bezpośrednio pochodzić z opisanego wyżej, probabilistycznego kroku związanego z szeregowaniem całego zbioru potencjalnych odbiorców. Podczas ewaluacji takiego podejścia zauważono jednak, że w takim wypadku końcowa wartość ulega częstym wahaniom prowadząc tym samym do dużego, niepożądanego zaszumienia wyjściowego sygnału (por. 2.3.2.1).

Wprowadzono dodatkowy moduł, którego celem jest wprowadzenie elementu zachowawczości do proponowanego rozwiązania. Moduł ten, modelujący problem jako *ukryty proces Markowa*, zapobiega zmianom wskazywanego obiektu tak długo, dopóki przesłanki o konieczności tejże nie staną się wystarczająco silne. W modelu tym *stanami* są fakty wskazywania poszczególnych obiektów, ich liczba jest więc równa liczności zbioru potencjalnych odbiorców uwagi. Korzystając z algorytmu zaproponowanego przez Viterbiego bazującego na idei programowania dynamicznego [For73], możliwe jest w każdej chwili wyliczenie sekwencji stanów, przez które przechodził system na podstawie danych wejściowych w zadanym oknie czasowym. Ostatni element takiej sekwencji może być traktowany jako aktualny, a więc wyjściowy omawianego systemu.

Aby uniknąć zbyt krótkiego czasu trwania rozpoznanej fiksacji w obrębie obiektu, prawdopodobieństwo pozostania we wcześniej wskazanym stanie zostało ustalone na znacznie większe, niż prawdopodobieństwo przejścia do innego stanu (w przypadku dalszych rozważań będą to odpowiednio wartości 0.95 i 0.05). Przekłada się to na zamodelowanie naturalnej, większej skłonności spojrzenia do śledzenia wciąż tego samego obiektu, niż przeskoczenia do innego.

Ciekawą cechą zastosowania ukrytego modelu Markowa jest jego zdolność do wskazania czasowej sekwencji stanów po nadejściu każdej z próbek w sposób niezależny. W ten sposób możliwe jest zrewidowanie poprzedniego wyboru po nadejściu nowych danych, jeśli wyjście z algorytmu jest wykorzystywane w trybie *off-line*. Z tego też powodu dla rozwiązań *on-line* opłacalne może się okazać wprowadzenie niewielkiego opóźnienia (rzędu 20 – 100ms dla częstotliwości próbkowania na poziomie 60Hz), odczytując wynikowe sekwencje dopiero po ustabilizowaniu wyjścia.

Sposób wykorzystania wartości wyjściowej zależy od konkretnego zastosowania śledzenia obiektu uwagi. Należy nadmienić, że w odróżnieniu od popularnie stosowanych metod detekcji fiksacji takich jak opisane wcześniej *I-DT* i *I-VT*, metoda *GDOT* dostarcza dyskretną informację o tym *który* obiekt, a nie *w którym miejscu* obiektu obserwator skupia uwagę. To z kolei rzutuje na podejście do konsumpcji rezultatów przez aplikację wykorzystującą moduł *GDOT*.

## 2.3.2 Ocena skuteczności identyfikacji

W celu wykazania skuteczności działania *GDOT* na tle powszechnie stosowanych algorytmów filtracji danych okulograficznych, konieczne było przeprowadzenie porównania jakości identyfikacji z użyciem poszczególnych technik. Należy się w tym miejscu zastanowić, jak można by zdefiniować samo pojęcie owej jakości, gdyż może się okazać, że nie jest to tak oczywiste.

W naiwnym ujęciu można by sądzić, że algorytm dobrze radzący sobie z poprawą jakości obarczonych błędem danych okulograficznych będzie najzwyczajniej dążył do minimalizacji udziału w czasie trwania całego badania tego czasu, w jakim wskazywany jest przezeń błędny obiekt. Jednak na podstawie wcześniejszych doświadczeń z pierwszymi próbami symulacji zjawiska głębi ostrości w oparciu o dane okulograficzne (por. 3.1) zauważono, że bardzo istotną przyczyną dyskomfortu użytkownika jest nie tyle sam udział błędu, co częstotliwość jego powstawania. Obserwatorzy bardzo często zgłaszali uciążliwość migotania obrazu, które powstawało w wyniku dostosowywania płaszczyzny ostrzenia do błędnie wyznaczonego obiektu, tym samym zmieniające zakres rozmycia poszczególnych elementów obrazu. Jeśli zmiany takie powstawały z dużą częstotliwością, rezultat był postrzegany jako bardziej błędny i niepożądany, niż gdy błędne wskazanie i co za tym idzie błędne rozmycie pozostawało niezmiennie przez dłuższy czas. Istotność problemu potęguje fakt, że algorytmy dążąc do minimalizacji udziału błędu wprowadzają bardzo częste korekty skutkujące zwiększeniem częstotliwości przeskoków pomiędzy wskazywanymi obiektami.

### 2.3.2.1 Metryka oceny jakości

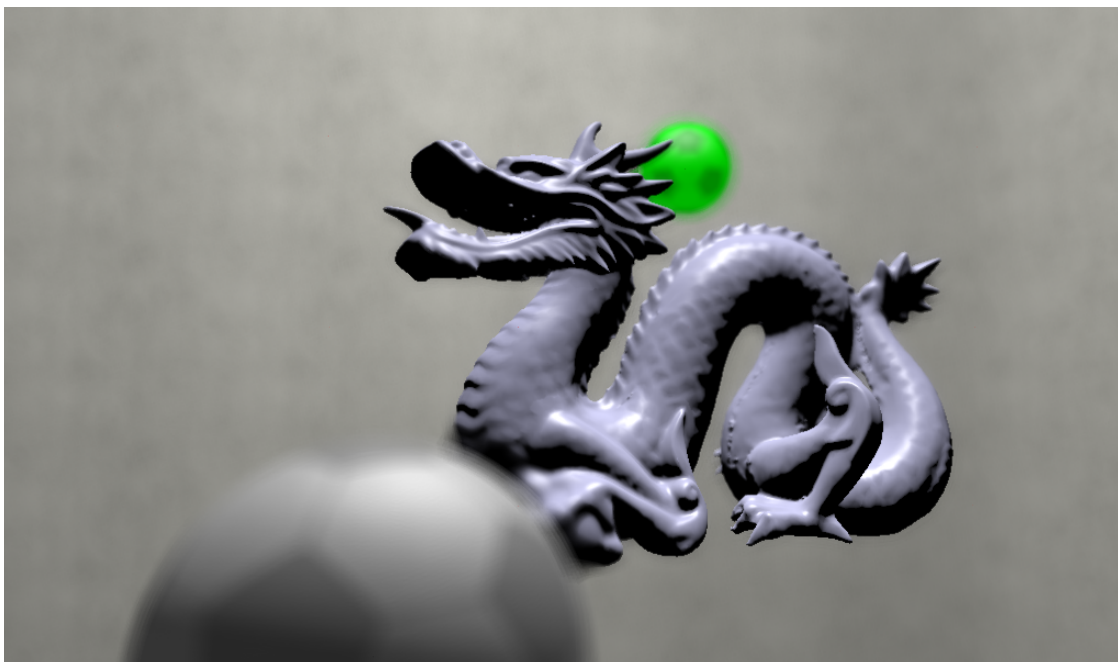
Do porównania skuteczności identyfikacji obiektu uwagi w oparciu o dane okulograficzne wykorzystano metrykę  $Q$ , na którą składają się dwa komponenty: *udział błędu identyfikacji*  $E_{rate}$  wyrażony jako stosunek czasu, przez który wskazywany jest błędny obiekt do całości czasu trwania sesji oraz *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.

Siedmiu uczestników zostało poproszonych o obejrzenie dziewięciu sekwencji wideo trwających po 18 sekund, syntezyzowanych w czasie rzeczywistym przez system komputerowy. Sekwencje przedstawiały wizualizację zielonej piłki obiegającej figurę smoka, przy czym obserwatorzy proszeni byli o śledzenie wzrokiem samej piłki (patrz Rys. 2.19). W każdej z sekwencji symulowane było zjawisko głębi ostrości z ostrością ustawioną na śledzoną piłkę, ale z imitowaną niedokładnością śledzenia o różnym stopniu  $E_{rate}$  oraz  $E_{freq}$ . Zadaniem uczestników było porównywanie ze sobą parami każdej z sytuacji poprzez wybranie tej, która w ich odczuciu była mniej drażniąca.

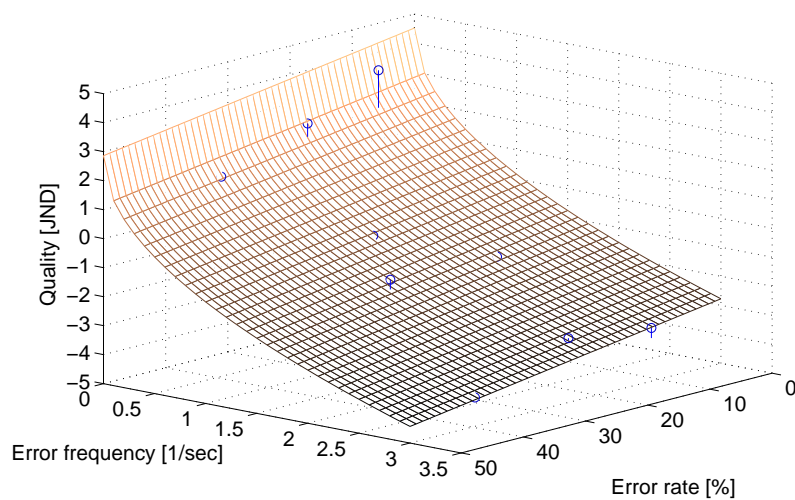
Otrzymano w ten sposób wyniki w postaci względnych ocen dla każdej z par pokrywających całą dwuwymiarową macierz porównań. By je przekształcić w spójną skalę mogącą stanowić narzędzie oceny jakości, przeniesiono je w jednowymiarową przestrzeń percepcyjnej oceny wyskalowaną w jednostkach *JND* (ang. *just noticeable*



Rysunek 2.19: Przykładowa klatka z animacji uwzględniającej wizualizację zjawiska głębi ostrości, która służyła do zebrania danych na potrzeby opracowania metryki jakości. Obserwator proszony był o śledzenie oznaczonego obiektu, w tym przypadku zielonej piłki.



Rysunek 2.20: Powierzchnia powstała w wyniku dopasowania do rezultatów uzyskanych z eksperymentu, który posłużył do zbudowania metryki jakości.





*difference*). Dokonano tego w oparciu o podejście Bayesowskie opisane w [SF01b]. Po dokonaniu aproksymacji uzyskanych rezultatów uzyskano równanie powierzchni prezentowanej na Rysunku 2.20:

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

Tak określona metryka faworyzuje sytuacje, w których częstotliwość zmian na błędne wskazanie jest najmniejsza. Wyliczając wartości  $E_{rate}$  i  $E_{freq}$  dla rezultatów uzyskanych poszczególnymi metodami identyfikacji można dokonać porównania wyników otrzymywanych za ich pomocą, uwzględniając przy tym odczucia wynikające z indywidualnych ocen obserwatorów dla przypadku, gdy rezultat związany jest z modyfikacją wyświetlanego obrazu w czasie rzeczywistym. Należy zaznaczyć, że metryka będzie prawidłowa dla oceny poprawności symulacji zjawiska głębi ostrości sterowanego okulografem. Dla innych zastosowań należałoby ponownie wykonać eksperymentalny dobór metryki. Pozwala ona jednak uzyskać wgląd w ogólne preferencje użytkowników.

### 2.3.2.2 Procedura eksperymentalna

Eksperyment oceny jakości działania metody *GDOT* polegał na prezentacji obserwatorom trzech animacji zszyntezowanych z użyciem technik grafiki komputerowej czasu rzeczywistego (patrz Rys. 2.21). 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. Poszczególnym animacjom, oznaczonym kolejno literami *A*, *B* i *C*, towarzyszyła wiedza o różnej liczbie rozmieszczonych na scenie obiektów będących potencjalnymi odbiorcami uwagi. Dla sekwencji *A* było ich 8, dla sekwencji *B* było 9, natomiast dla najbardziej złożonej sekwencji *C* zdefiniowano 110 takich obiektów. Należy przy tym wspomnieć, że ta ostatnia była najbardziej zbliżona do sytuacji możliwych do spotkania w rzeczywistych aplikacjach okulografii czasu rzeczywistego związanych z wirtualną rzeczywistością.

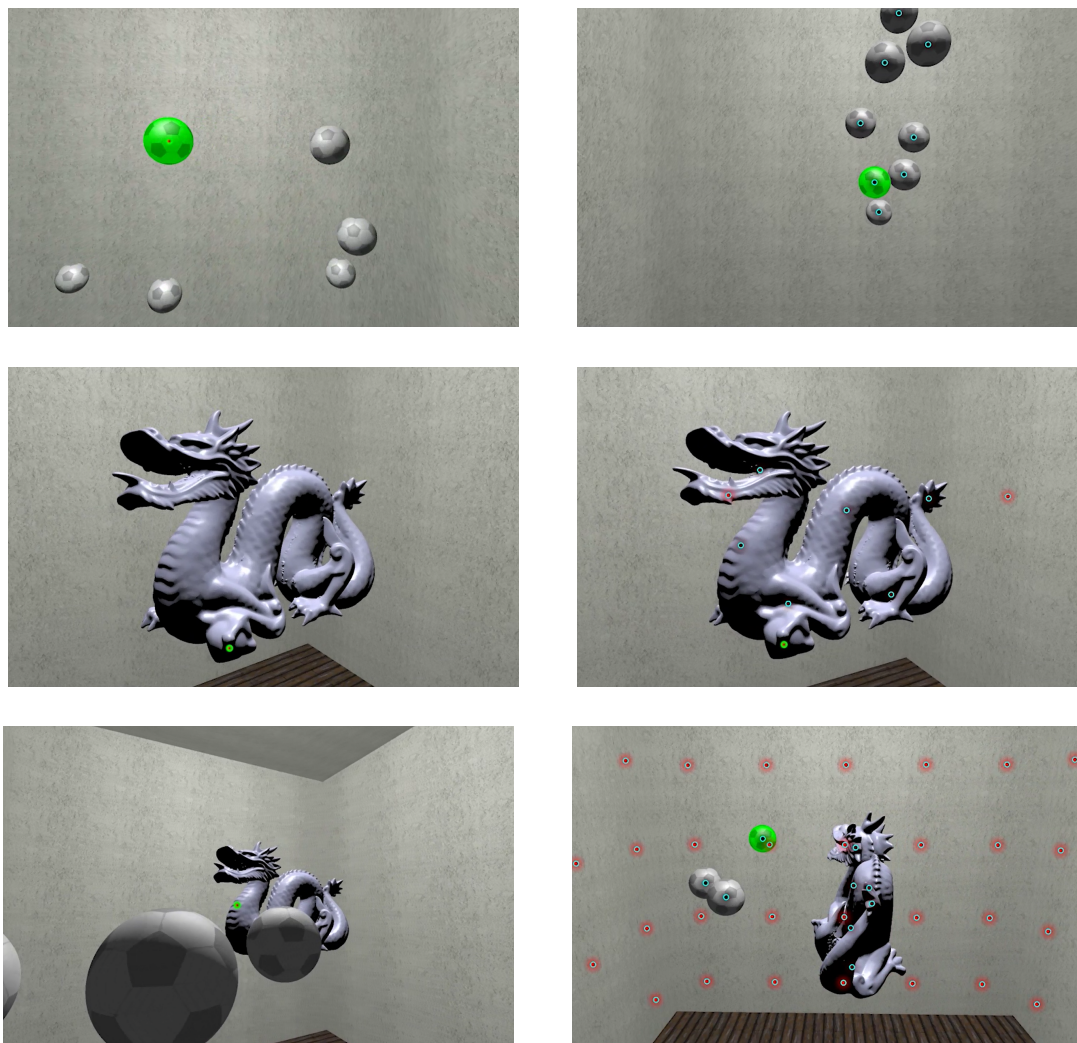
Przez cały czas ekspozycji sekwencji wideo na ekranie zaznaczony był punkt, o dokładne śledzenie którego proszono uczestników. Punkt ten podążał za poszczególnymi obiektami lub ich fragmentami oraz przemieszczał się pomiędzy nimi, mając za zadanie imitować ukierunkowanie uwagi osoby oglądającej scenę. W eksperymencie założono, że obserwator patrzy na ten punkt, a więc to właśnie położenie tego punktu powinno być identyfikowane przez algorytm detekcji fiksacji.

Po zakończeniu sesji zebrane dane poddano filtracji z użyciem różnych metod identyfikacji uwagi wizualnej. Zasymlowano działanie tych metod w czasie rzeczywistym, a więc nie przewidziano możliwości odwoływania się do sygnału przyszłego.

W badaniu wzięło udział 39 osób. Obraz prezentowany był na 22-calowym ekranie oddalonym o 65 centymetrów od obserwatora, przy czym odległość ta była wymuszona poprzez zastosowanie podbródka stabilizującego położenie głowy.

### 2.3.2.3 Kalibracja parametrów

Jakość identyfikacji obiektów z użyciem *GDOT* porównano z metodami *I-DT*, *I-VT* oraz wybieraniem obiektu na podstawie surowych danych pochodzących z oku-



Rysunek 2.21: Przykładowe klatki z animacji użytych podczas procedury eksperymentalnej. Kolejno od góry, sceny: *A*, *B*, *C*. Pierwsza kolumna pokazuje klatki widoczne przez obserwatora, druga kolumna zawiera wizualizację zbioru potencjalnych odbiorców uwagi.

lografu.

Wartości parametrów wszystkich metod identyfikacji, które zostały poddane porównaniu, zostały zoptymalizowane dla środowiska badań. Optymalizacji dokonano z użyciem operującej na ciągłych wartościach, zbieżnej metody simpleks (ang. *downhill simplex method*), dzieląc dane zebrane dla scen na dwa zbiory: treningowy (uczący) oraz testowy. Porównania dokonano z użyciem wartości otrzymanych przy użyciu tego drugiego. Optymalizacja została dokonana zarówno dla poszczególnych sekwencji, jak i dobrano zestaw wartości parametrów, który najlepiej sprawdzał się dla wszystkich z nich. Optymalizowane parametry wraz z uzyskanymi wartościami zebrano w Tabeli 2.2.

Tabela 2.2: Zoptymalizowane wartości parametrów podczas porównywania metod identyfikacji obiektów.

Metoda	Parametr	Optymalizacja dla:			
		A	B	C	ABC
I-VT	<i>velocity</i> [°/s]	3.22	5.90	4.76	4.02
I-DT	<i>duration</i> [s]	175	181	200	181
	<i>dispersion</i> [piksele]	245	173	178	208
	<i>window length</i> [s]	45.3	18.5	11.4	20.1
GDOT	$\sigma_s$ [piksele]	469	244	443	465
	$\sigma_v$	0.65	0.68	0.38	0.73
	$\omega_p$	1	1	1	1
	$\omega_v$	0.54	0.39	0.70	0.41

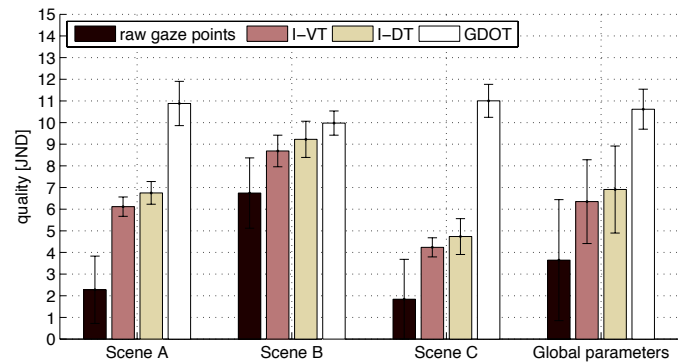
Ważnymi parametrami GDOT są cechy dolnoprzepustowej filtracji prędkości próbek oraz opóźnienie odczytu wyników działania *HMM*. Są to parametry dyskretne i z tego powodu nie można ich włączyć do użytej procedury optymalizacyjnej. Niemniej, jak wykazała analiza wyników eksperymentów, korelacja pomiędzy tymi, a pozostałymi parametrami jest relatywnie niewielka. To pozwoliło na poszukiwanie najlepszych wartości także i dla tych parametrów w oddzielnym kroku, w wyniku czego wyznaczono szerokość okna filtracji prędkości na 120 próbek (2 sekundy przy założeniu częstotliwości aktualizacji na poziomie 60 Hz), a opóźnienie odczytu *HMM* na 18 próbek (300ms dla 60 Hz).

### 2.3.2.4 Rezultaty

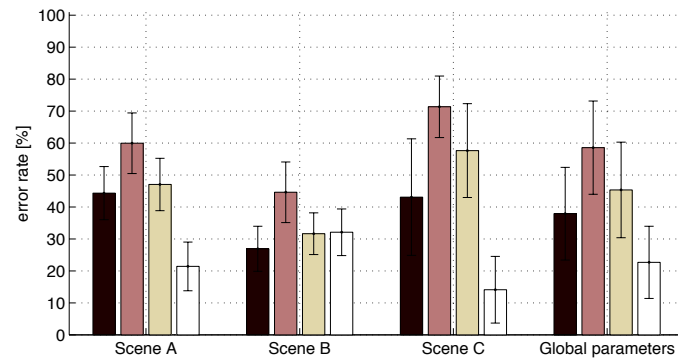
Dla testowego zbioru danych uzyskano wyniki, które wskazują na znacznie lepszą jakość identyfikacji obiektów z użyciem metody *GDOT*, niż uzyskuje się korzystając z innych metod (zob. Rys. 2.22a).

Warto zwrócić uwagę, że przewaga metody *GDOT* jest mniej znacząca dla animacji *B*. Jest to przykład, w którym na ekranie było niewiele poruszających się elementów.

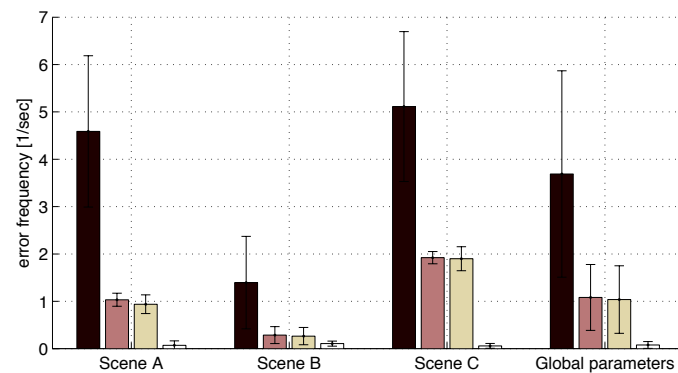
Surowe dane okulograficzne, nie poddane filtracji, charakteryzują się stosunkowo niewielkim udziałem błędów w czasie w porównaniu z pozostałymi przypadkami. Jednocześnie częstotliwość błędnej identyfikacji wielokrotnie przewyższa rezultaty



(a)



(b)



(c)

Rysunek 2.22: 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*.

działania metod filtracji. To ostatnie jest związane z dużym zaszumieniem sygnału, a więc brakiem precyzji samych współrzędnych punktu skupienia wzroku, które zwraca system okulograficzny. Można przypuszczać, że z racji przestrzennego rozrzutu kolejnych próbek w czasie, niektóre z nich nie trafiają w obiekt na którym obserwator skupia swoją uwagę. Wystarcza to jednak, by takie nietrafienie zostało zaliczone jako uwzględnienie innego obiektu i doprowadziło do błędnej identyfikacji, o ile sygnał nie zostanie uprzednio poddany żadnej filtracji.

Dla przypadków związanych z metodami  $I-VT$  oraz  $I-DT$  obserwujemy, zgodnie z oczekiwaniami, znaczną redukcję częstotliwości błędu poprzez zmniejszenie zaszumienia współrzędnych próbek kolejno zwracanych przez system. Nie odbywa się to jednak bez wpływu na ogólny udział błędu, gdzie da się zauważyć obniżenie jakości otrzymywanych rezultatów. Przyczyny można dopatrywać się w powstającym w wyniku filtracji opóźnieniu oraz w przesuwaniu przefiltrowanego punktu skupienia wzroku w wyniku błędy systematycznego danych, co będzie lepiej widoczne na Rys. 2.23.

Cechą szczególną metody  $GDOT$  jest natomiast znaczne ograniczenie częstotliwości przechodzenia do wskazywania błędnego obiektu, co jest wynikiem zachowawczości algorytmu płynącej z użycia na jego wyjściu warstwy opartej na ukrytym modelu Markowa. 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. 2.22a.

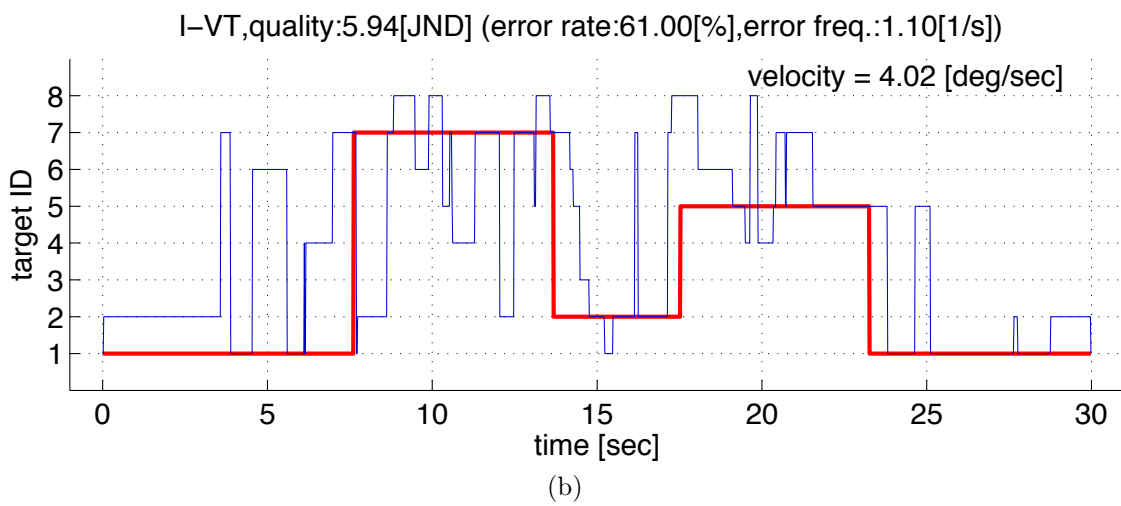
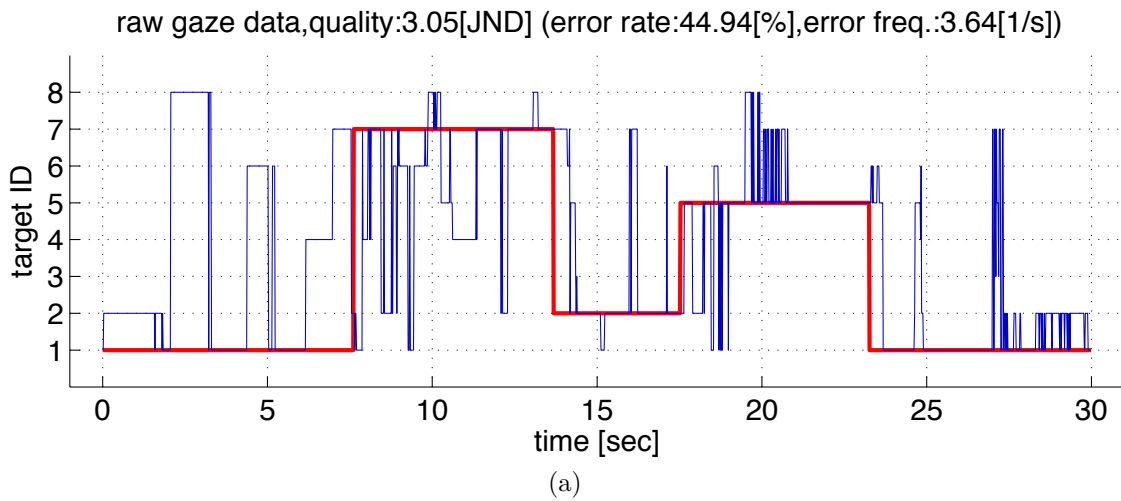
Na Rys. 2.23 przedstawiony jest przykładowy przebieg czasowy identyfikacji obiektów uwagi dla różnych metod filtracji danych. W idealnym przypadku linia niebieska, będąca zobrazowaniem wyjścia identyfikacji obiektu uwagi, powinna pokrywać się z linią czerwoną oznaczającą ten obiekt, którego śledzenie było zadaniem obserwatora.

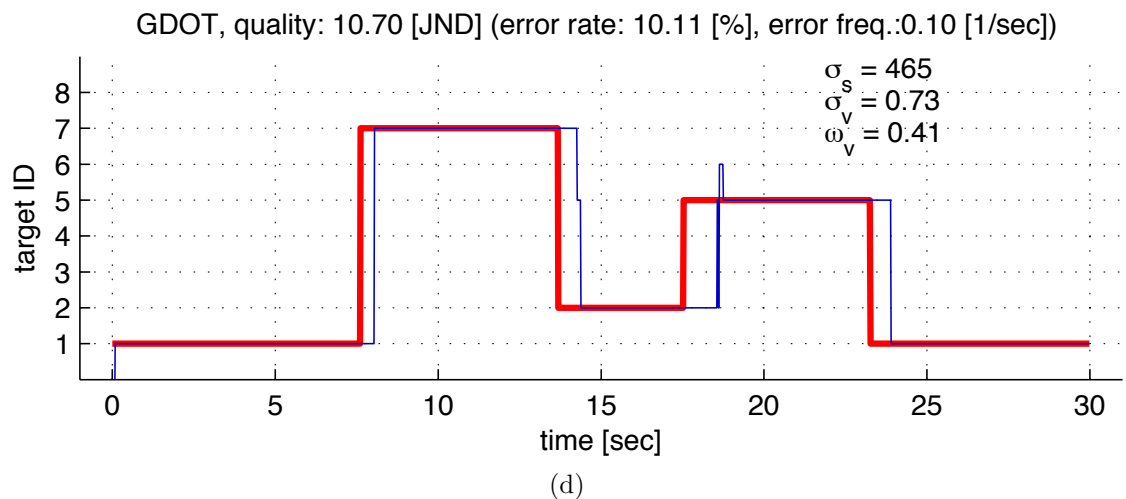
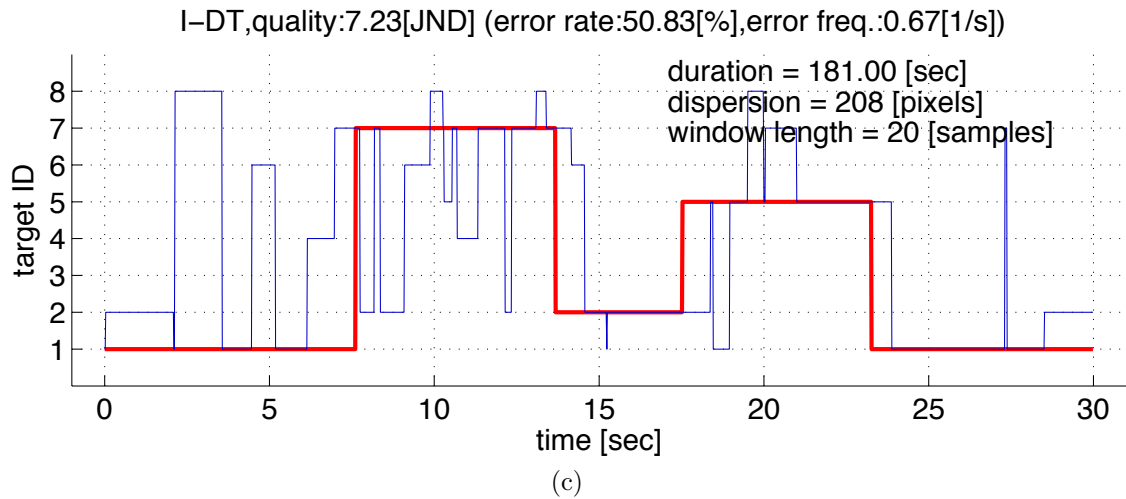
Wykres dla surowych danych charakteryzuje się największą zmiennością. Jest to więc zgodne z wcześniejszą obserwacją dotyczącą częstotliwością błędu identyfikacji. W wyniku dużego zaszumienia sygnału wejściowego, na wyjściu również otrzymujemy niestabilny rezultat.

Użycie metod filtracji takich jak  $I-VT$  czy  $I-DT$  skutkuje redukcją częstotliwości zmian. Da się jednak zauważyć wprowadzenie nieznacznego opóźnienia względem sygnału referencyjnego oraz tendencję do podtrzymywania błędnej identyfikacji, które skutkują zwiększonym udziałem błędu w całości czasu obserwacji.

Metoda  $GDOT$  najlepiej odwzorowuje sygnał referencyjny. Niewielki udział 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.





Rysunek 2.23: 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].

## 2.4 Zastosowania metody GDOT

Należy nadmienić, że zakres stosowalności *GDOT* nie ogranicza się jedynie do danych okulograficznych. Podejście to może być wykorzystywane w każdej sytuacji, gdy celem jest wyznaczenie ze zbioru potencjalnych kandydatów o znanym,  $n$ -wymiarowym położeniu tego, który z największym prawdopodobieństwem jest śledzony za pomocą zaszumionego i obciążonego błędem sygnału określonego w tej samej przestrzeni. Niemniej jednak warto zwrócić uwagę na charakterystyczne przykłady użycia *GDOT* w dziedzinie *eye trackingu*, które zostały zaproponowane, opracowane i przebadane w trakcie opracowywania niniejszej rozprawy.

W tym podrozdziale omówione zostaną trzy sposoby wykorzystania metody *GDOT*: identyfikacja przedmiotu zainteresowania na potrzeby analizy uwagi osób oglądających wideo (Rozdział 2.4.1, [BM14a]), przyspieszenie syntezy obrazu poprzez redukcję jego szczegółowości w częściach nieobjętych szczegółowym widzeniem użytkownika (Rozdział 2.4.2, [TB10]) i projekt rozszerzeń klasycznych elementów interfejsu użytkownika w grze komputerowej, opartych o wiedzę o oglądanym w danej chwili obiekcie (Rozdział 2.4.3, [BM14b]).

### 2.4.1 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. Zagadnienie to jest nie tylko istotne z punktu widzenia badań z dziedziny psychologii poznawczej, ale wzbudza też zainteresowanie ze strony badań marketingowych. Łatwo dostrzec potencjał w wiedzy o tym jakie elementy widoczne na ekranie przyciągnęły uwagę widza podczas wyświetlania klipu promującego produkt albo jak zauważalne i skuteczne jest zastosowanie techniki lokowania marki w danym filmie fabularnym.

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.

Należy zwrócić uwagę, że obecnie najczęściej podczas analizy danych okulograficznych zebranych podczas ekspozycji bodźca będącego obrazem wideo, dane te poddawane są klasyfikacji związanej z wyodrębnieniem kolejnych, statycznych fiksacji [Voß09, Sen10]. Jest to jednak sprzeczne z naturą takiego obrazu, który zwykle zawiera ruchome elementy, a które nie są śledzone przez widza z użyciem sekwencji fiksacji, lecz za pomocą płynnego pościgu. Tym samym filtracja danych okulograficznych powinna uwzględniać ten fakt, by uniknąć wprowadzania dodatkowej niedokładności jaka może się pojawić gdy taki płynny pościg jest błędnie interpretowany jako ciąg krótkich fiksacji rozdzielonych sakadami.

Kolejny kłopot z badaniem eyetrackingowym na obrazach wideo polega na konieczności powiązania ze sobą obrazów danych obiektów w kolejnych klatkach. O ile w przypadku statycznych obrazów sprowadza się to najczęściej do ręcznego wy-



znaczenia obszaru zainteresowania dla każdego z potencjalnie interesujących elementów obrazu, to w przypadku wideo dodatkowo konieczne jest wskazanie toru przemieszczania się takich obiektów w czasie oraz ewentualnych jego transformacji. W popularnym na rynku oprogramowaniu służącym do wspomagania tego rodzaju analiz [Voß09] najczęściej wiąże się to z mozolnym i trudnym dla użytkownika oznaczaniem każdej klatki osobno.

Problematyczność obu opisanych wyżej aspektów może być zredukowana z pomocą algorytmu *GDOT* oraz jego specyficznego wzbogacenia o moduł współdziałający ze śledzeniem w oparciu o technikę przepływu optycznego (ang. *optical flow*). Jest to autorskie rozwiązanie powstałe w ramach pracy nad niniejszą rozprawą i opublikowane w ramach artykułu [BM14a].

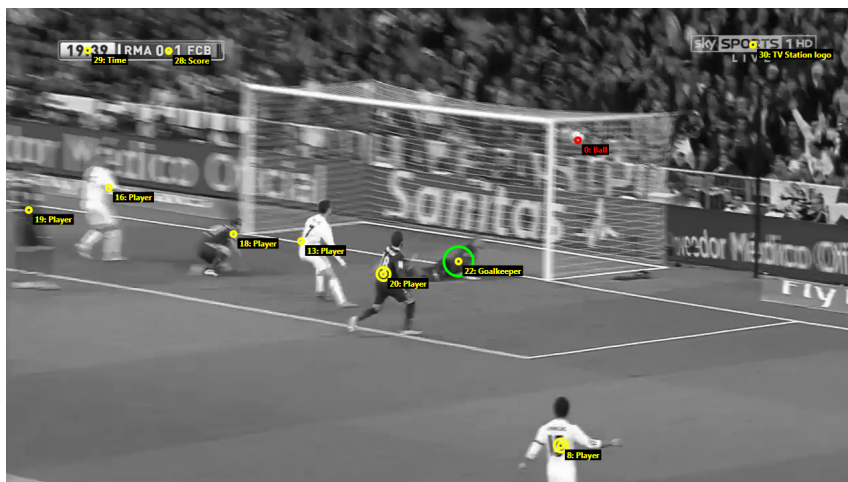
#### 2.4.1.1 Wykorzystanie przepływu optycznego

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 położenia wybranych przez siebie elementów we wszystkich kolejnych klatkach sekwencji wideo. W zamian wykorzystywana jest technika analizy rzadkiego (ang. *sparse*) przepływu optycznego, która pozwala w sposób zautomatyzowany śledzić poruszanie się fragmentu sceny w czasie.

Wyznaczone przez eksperta obiekty traktowane są jako singletony w przestrzeni obrazu, posiadają więc dwuwymiarowe współrzędne określające pozycję punktu, który został wskazany. Z punktem tym naturalnie związany jest też piksel obrazu, dla którego z użyciem metody Lucasa-Kanade [LK81] wyznaczany jest przepływ optyczny pomiędzy kolejnymi klatkami. W ten sposób możliwe jest śledzenie ruchu całego obiektu na podstawie lokalnego sąsiedztwa tego punktu, o ile tylko jako obraz nie ulegnie znaczącej zmianie, nie zostanie on przesłonięty lub nie znajdzie się poza obszarem ekranu. Dlatego też oprócz metody zautomatyzowanej, oprogramowanie wspomagające wyznaczanie zbioru potencjalnych odbiorów uwagi, umożliwi także ręczne korygowanie wyznaczonego toru z wykorzystaniem pomocnych narzędzi, jak na przykład liniowa interpolacja w czasie pomiędzy dwoma wskazanymi punktami. Jest to szczególnie ważne z uwagi na fakt, że rzeczywiste obrazy wideo często posiadają zniekształcenia w postaci szumu, niekonsekwentnego i niepłynnego ruchu, rozmycia ruchu, gdy czas naświetlania kolejnych klatek jest długi (ang. *motion blur*). Takie zjawiska niekorzystnie wpływają na skuteczność działania wyznaczania przepływu optycznego [HFC05], więc ręczna korekta przez użytkownika jest w niektórych sytuacjach niezbędna. Szczególnie, że chwilowe przesłonięcie obiektu który jest obserwowany przez widza nie powoduje przerwania płynnego pościgu [BF85], co z punktu widzenia efektywności zastosowania metody *GDOT* ma szczególne znaczenie.

Jako element dalszych prac warto rozważyć podjęcie próby całkowitego wyłączenia udziału eksperta w tworzeniu zbioru obiektów. Poprzez zastosowanie algorytmów z dziedziny zautomatyzowanej detekcji zauważalności elementów obrazu (ang. *saliency detection*), możliwe byłoby wyznaczenie obszarów obrazu, które mają dużą

Rysunek 2.24: 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.



szansę stać się potencjalnym odbiorcą uwagi. Jest to jednak dość optymistyczne założenie spłycające rolę semantycznej wartości elementów obrazu jedynie do wizualnie wyróżniających się fragmentów, którego skuteczność należałoby poddać odrębnemu badaniu. Możliwe jednak, że tak uzyskany wstępny rezultat mógłby posłużyć za punkt wyjścia, znacząco upraszczający zadanie stawiane przed osobą przygotowującą analizę.

#### 2.4.1.2 Eksperyment

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 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. 2.24).

W badaniu użyto trzech krótkich sekwencji (trwających odpowiednio 18, 29 i 13 sekund), nazwanych kolejno *A*, *B* i *C*. Oglądaniu każdej z tych sekwencji odpowiadało inne polecenie. W przypadku *A*, obserwatorzy mieli za zadanie oglądać mecz w sposób całkowicie swobodny, bez narzuconych żadnych ograniczeń (ang. *free examination*). Podczas ekspozycji fragmentu *B* wyświetlano na ekranie kolorowy

Rysunek 2.25: Marker widoczny podczas ekspozycji sekwencji *B*. Zadaniem obserwatora było śledzenie jego położenia za pomocą spojrzenia.



marker, a oglądających proszono o uważne jego śledzenie (por. Rys. 2.25). Marker ten wskazywał kolejno różne obiekty na ekranie, wymuszając przenoszenie uwagi obserwatora. W ten sposób uzyskano przebieg mający być referencją dla oceny skuteczności działania badanej metody. W przypadku *C*, zadaniem oglądających osób było natomiast śledzenie piłki, co również pozwoliło na porównanie zarejestrowanego i poddanego filtracji przebiegu punktu skupienia wzroku z przebiegiem referencyjnym. Cały eksperyment był powtórzony trzykrotnie dla każdego z obserwatorów, a podczas każdego powtórzenia używano okulograf do zbierania danych.

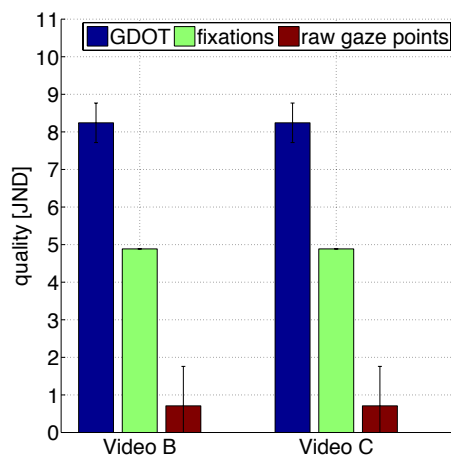
Uzyskane przebiegi punktów skupienia wzroku poddano działaniu detekcji fiksacji w oparciu o powszechnie stosowaną metodę *I-DT*. Porównano wyniki identyfikacji obiektów na podstawie tak pozyskanych danych dla trzech przypadków: gdy źródłem były surowe dane, gdy identyfikację oparto na fiksacjach oraz gdy za identyfikację w całości odpowiadał algorytm *GDOT* (zob. Rys 2.27). 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. 2.26.

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*.

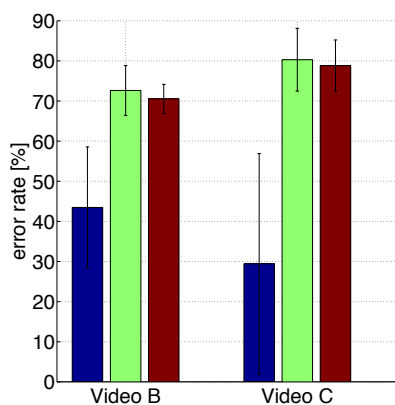
Oznacza to więc, że zastosowanie techniki *GDOT* nie jest ograniczone wyłącznie do obrazów powstających w wyniku komputerowej syntezy. Także w innych sytuacjach, gdy jesteśmy w stanie wyróżnić zbiór rozróżnialnych obiektów na prezentowanym obserwatorowi obrazie, możliwe jest wykorzystanie tej techniki do identyfikacji tego z obiektów, na który jest skierowany wzrok użytkownika.

## 2.4.2 Przyspieszenie syntezy obrazu

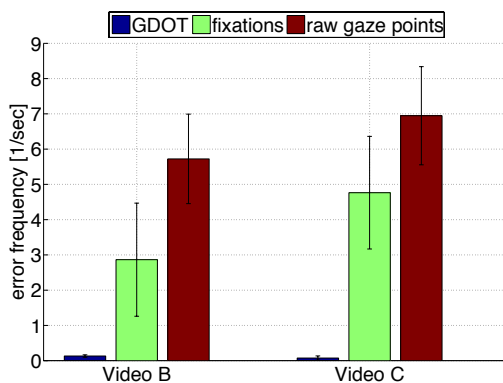
Obliczenia związane ze złożonymi efektami graficznymi stosowanymi do już zsyntezowanych klatek (ang. *post processing*) wymagają dużej ilości zasobów. Tego ro-



(a)

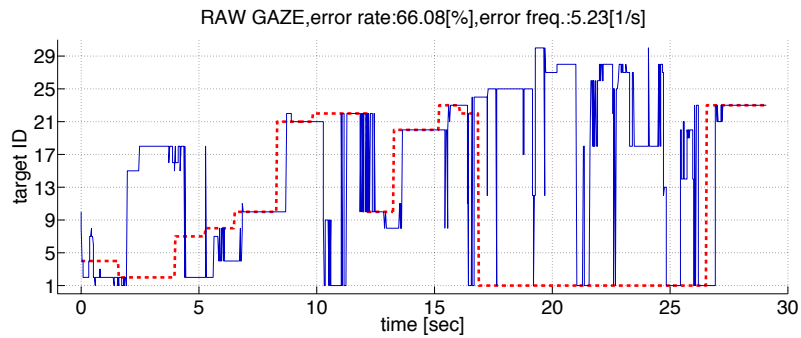


(b)

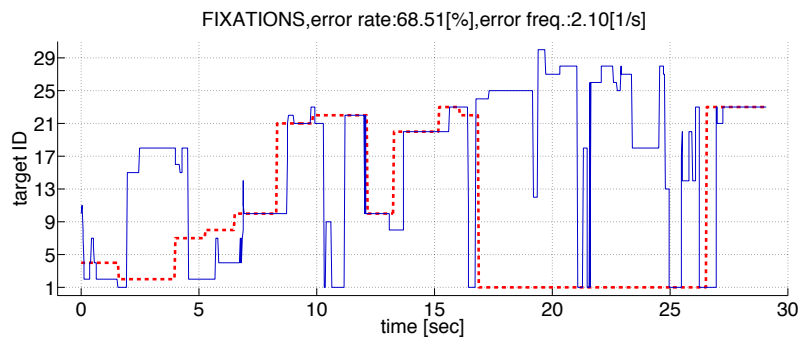


(c)

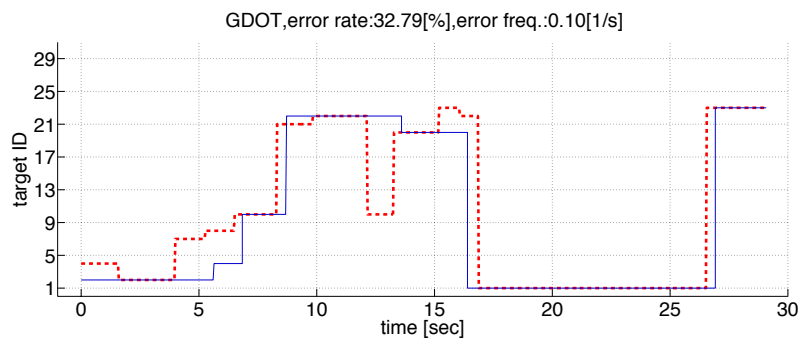
Rysunek 2.26: Jakość identyfikacji według opisanej w rozprawie metryki (a) oraz dwie składowe: udział błędu (b) i jego częstotliwość (c), z podziałem na poszczególne sekwencje *B* i *C*. Ilustracja zaczerpnięta z autorskiej publikacji [BM14a].



(a)



(b)



(c)

Rysunek 2.27: Przykładowe przebiegi czasowe identyfikacji obiektu uwagi w sekwencji  $B$  z użyciem różnych metod filtracji, uzyskane dla tego samego zestawu danych okulograficznych 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.

dzaju 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 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 (por. Rys. 1.4).

#### 2.4.2.1 Obszar zainteresowania (*ROI*)

Korzystając z systemu okulograficznego, możliwe jest wskazanie chwilowego obszaru na ekranie, który jest pokrywany widzeniem szczegółowym. Tym samym uwzględniając tę wiedzę, możliwe jest ograniczenie rozmiaru części obrazu, na której prezentowany jest kosztowny w syntezie obraz wysokiej jakości. Pozostała część obrazu może zawierać wówczas obraz o obniżonej jakości. Przy czym obniżenie to nie powinno zmieniać ogólnego charakteru obrazu, jego kolorystyki, niskiej częstotliwości kontrastu lub innych cech, które mogłyby zostać wychwycone nawet przez widzenie peryferyjne [GFD<sup>+</sup>12]. Zasadne wydaje się tutaj posłużenie czułością na kontrast jako funkcją odchylenia od kierunku widzenia (ang. *contrast sensitivity as a function of eccentricity*) [RFN92], aczkolwiek zauważalność zmian związanych z pogorszeniem jakości w polu peryferyjnym nie zależy wyłącznie od kontrastu.

Optymalny rozmiar *ROI* jest więc ograniczony z dwóch stron. Ograniczeniem z dołu jest minimalna wielkość odpowiadająca wyraźnemu widzeniu przez człowieka tej cechy obrazu, której jakość ma zostać pogorszona. Górne ograniczenie natomiast wynika z powiązania wielkości części obrazu mającej zostać wyrenderowana w wysokiej jakości z ilością zasobów, które są konieczne do dokonania syntezy tego obszaru. Można przyjąć, że im rozmiar *ROI* jest mniejszy, tym większy będzie zysk wydajności systemu syntezującego obraz.

*ROI*, które miałyby dokładnie pokrywać obszar widziany szczegółowo, powinno mieć kształt zbliżony do poziomej elipsy (por. [PWBoNR73], str. 631). Jednak z dwu powodów takie podejście nie jest optymalne. Po pierwsze, należy pamiętać o niedoskonałości danych otrzymywanych z systemu okulograficznego. Trudno więc zakładać, że środek *ROI* będzie się znajdował w faktycznym miejscu, gdzie znajduje się w danej chwili punkt skupienia wzroku. Tym samym konieczne jest wzięcie tego pod uwagę i wprowadzenie marginesu tolerancji, który redukuje negatywny skutek ewentualnych przesunięć. Po drugie, wykonanie sprawdzenia czy przetwarza-

ny w danej chwili piksel (zakładając, że przyspieszany element potoku renderowania dotyczy efektu graficznego który jest obliczany dla każdego piksela bufora klatki osobno) należy do elipsy bądź bardziej skomplikowanego kształtu, jest niepotrzebnie kosztowne. Połączenie ze sobą obu tych problemów prowadzi do konkluzji, że niejednokrotnie rozsądnie jest uznać za przybliżenie *ROI* prostokąt, który da się też wydajniej poddać kompozycji podczas obliczania finalnej, pełnoekranowej klatki do wyświetlenia.

#### 2.4.2.2 Zastosowania

Oprócz zastosowań związanych z efektami graficznymi opartymi na działaniach na buforze klatki, w literaturze opisano także próby zastosowania tej techniki dla innych kosztownych zadań syntezy obrazu. Przykładem jest adaptacyjna szczegółowość geometrii trójwymiarowych brył [Duc07] oraz rendering w oparciu o metodę śledzenia promieni (ang. *ray tracing*) [Sie14, RWH<sup>+</sup>16]. W obu tych przypadkach, obliczenia dotyczące geometrii sceny są znacznie bardziej szczegółowe w wycinku przestrzeni sceny odpowiadającym fragmentowi ekranu wyznaczonemu przez *ROI*.

Swojego rodzaju złączeniem obu charakterów przyspieszanych operacji jest metoda adaptacyjnej wersji oświetlenia metodą *ambient occlusion* [MJ12]. Liczba odbitych promieni dla pikseli znajdujących się w obszarze *ROI* jest znacząco większa, niż poza nim. Autorzy zmagali się jednak z problemem zauważania przez widza zmian występujących przy granicy poruszającego się obszaru wysokiej jakości, związanego z estymowanym na podstawie danych okولوجraficznych punktem skupienia wzroku.

### ROI w symulacji zjawiska głębi ostrości

Zaproponowano próbę przyspieszenia symulacji zjawiska głębi ostrości, która odbywa się metodą opisaną dalej w niniejszej rozprawie (zob. Rozdz. 3.1). To autorskie podejście zostało opublikowane w artykule [TB10].

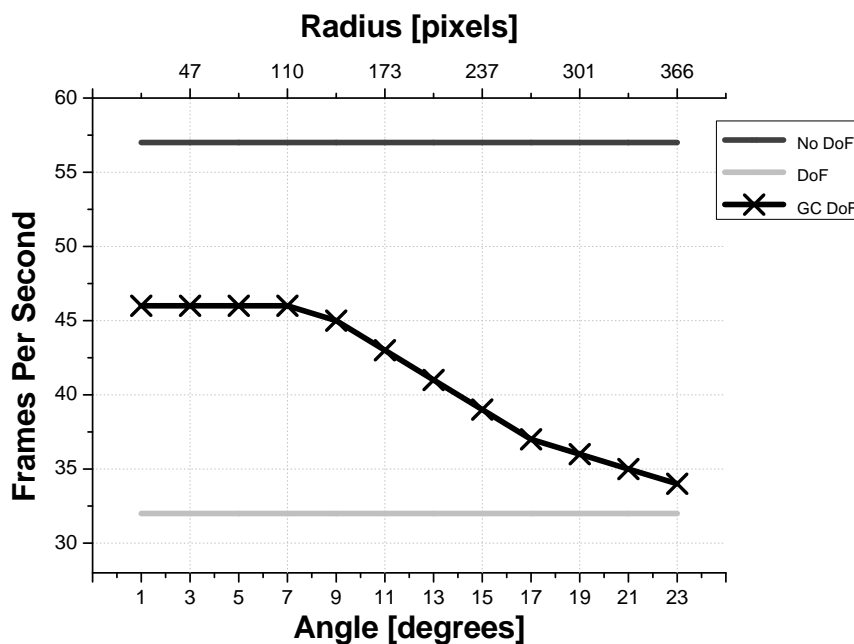
Przyjęto, że obszar poza *ROI* będzie renderowany do bufora klatki o rozdzielczości dwukrotnie mniejszej wzdłuż każdego z boków niż docelowa rozdzielczość finalnej klatki. Zawartość bufora o pomniejszonej rozdzielczości będzie następnie wyświetlana na ekranie z użyciem filtracji dwuliniowej. Tym samym 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  $800 \times 600$  pikseli (zob. Rys. 2.28). Rezultaty przedstawiono na Rysunku 2.29. Okazuje się, że zakładając kątowny rozmiar *ROI* z punktu widzenia obserwatora na poziomie  $15^\circ$ , 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.

Rysunek 2.28: 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 2.29: 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]





### 2.4.2.3 Problemy i dalsze prace

Kluczowym problemem związanym z przyspieszaniem pracy systemu syntezującego obraz poprzez obniżenie jakości rezultatów w obszarach nieobjętych widzeniem centralnym, jest wyeliminowanie zauważania przez użytkownika obszarów o obniżonej jakości. Nie jest to jednak kwestia wyłącznie odpowiedniego doboru rozmiaru *ROI* tak, aby odpowiadał obszarowi wyraźnego widzenia. Przyczyną problemu jest również opóźnienie systemu okulograficznego, który dostarcza dane o aktualnym położeniu punktu skupienia wzroku, a także jakość tych danych. Duże, zauważalne opóźnienie prowadzi do sytuacji w których po przeniesieniu wzroku w inną część ekranu, użytkownik zdąży zauważyć obraz o pogorszonej jakości [CM16].

Niska dokładność i błąd precyzji estymowanego punktu skupienia wzroku, mogą z kolei prowadzić do przesunięcia obszaru *ROI* względem faktycznego miejsca, które jest oglądane. Kolejnym niepożądanym skutkiem szumu jest przemieszczanie się *ROI* ze stosunkowo dużą częstotliwością, co może dodatkowo ułatwić zauważenie krawędzi obszaru które w przypadku braku poruszania się *ROI* byłyby nieruchomym obrazem, lub obrazem pozbawionym tak charakterystycznego ruchu. Na niekorzyść działa tutaj stosunkowo wysoka czułość aparatu widzenia człowieka na ruch w peryferyjnych częściach pola widzenia, której degradacja wraz ze zwiększającym się kątem widzenia jest wolniejsza niż w przypadku czułości na kontrast [MN84, RFN92].

Zauważalność krawędzi można też zredukować stosując płynne mieszanie (ang. *blending*) *ROI* wysokiej jakości z niskiej jakości pełną klatką, wraz z oddalaniem się od pomierzonego punktu skupienia wzroku. Należy jednak zwrócić uwagę na to, że w takiej sytuacji konieczne może być zdublowane obliczenie każdego z pikseli też w niższej jakości, także tych pikseli które pokrywają się z samym *ROI*. To, jak i istotna kosztowność operacji ważonego łączenia wartości pikseli pochodzących z dwu buforów, może stanowić o złożoności takiego ulepszenia.

Problematyczność opóźnienia w wyświetlaniu obrazu dobrej jakości po przeniesieniu wzroku w część, która do tej pory była uznawana za znajdującą się w peryferyjnej części pola widzenia, można zminimalizować jedynie dążąc do redukcji czasu potrzebnego na estymację punktu skupienia wzroku lub wprowadzając pewnego rodzaju predykcję potencjalnych ruchów. Z kolei problem wpływu niskiej dokładności i precyzji na niepełne obejmowanie rzeczywistego obszaru widzenia centralnego lub na wysokoczęstotliwościowe zmiany jego położenia, można próbować rozwiązać poprzez zwiększenie rozmiaru *ROI* z uwzględnieniem zakładanego błędu dokładności, jak i przez wprowadzenie filtracji położenia. Filtracja taka może bazować na detekcji fiksacji lub metodzie identyfikacji obiektu, na którym skupiana jest uwaga obserwatora (por. Rozdział 2.2.4). Należy jednak pamiętać, że w celu eliminacji niepożądanego zjawiska występowania zmian w obrazie, który w innym wypadku byłby statyczny, położenie *ROI* powinno zmieniać się jak najrzadziej.

Istotnym czynnikiem który należy rozpatrywać jest fizyczny rozmiar prezentowanego obrazu. Wyświetlając obraz na ekranie o bardzo dużej przekątnej, rozmiar *ROI* rozumiany jako część pełnej klatki, mógłby być znacznie mniejszy niż przy małym ekranie.

### 2.4.3 Interfejsy użytkownika oparte na danych okulograficznych

Osobną gałęzią zastosowań okulografii 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.

Poniżej przedstawiono przegląd opisanych w literaturze rozwiązań, dalej zaś zaproponowano i przebadano skuteczność autorskich, bazujących na wykorzystaniu kierunku spojrzenia, form interakcji z wirtualną rzeczywistością.

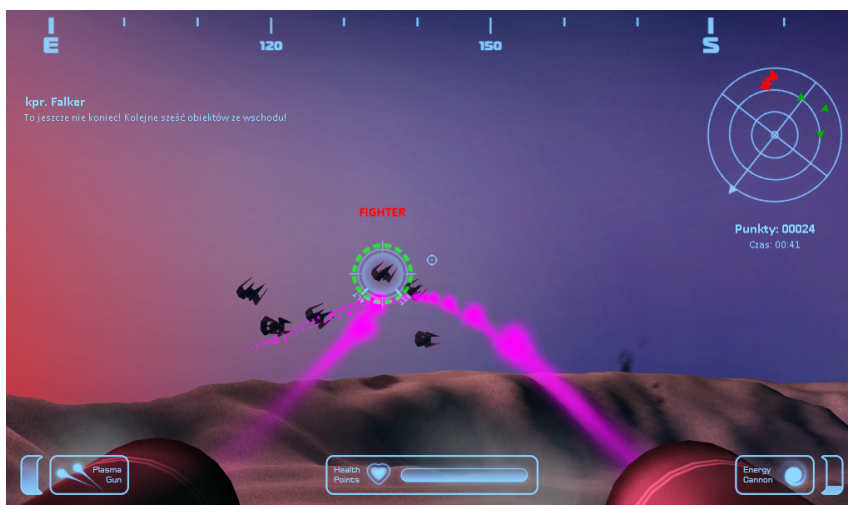
#### 2.4.3.1 Interakcja za pomocą spojrzenia

Najprostszą formą wykorzystania zbieranych na bieżąco danych okulograficznych podczas swobodnej nawigacji w wirtualnym środowisku, jest pośrednie powiązanie ruchu kamery ze zmianami kierunku widzenia. Podstawowe rozwiązanie w tym zakresie zostało opisane przez Smitha i Grahama [SG06], gdzie autorzy zmodyfikowali powszechnie znaną grę komputerową tak, by możliwa była kontrola prowadzonej przez gracza postaci z użyciem wzroku. Kamera jest zawsze obracana w taki sposób, by obiekt na który spogląda gracz przemieszczał się w centralną część ekranu. Subiektywne odczucia ankietowanych osób wskazały jednak podatność takiego podejścia na tzw. problem dotyku Midasa (por. [Jac91]), a więc zjawisko niechcianej interakcji odbywającej się bez intencji ze strony użytkownika. Podobne rozwiązanie zostało zaproponowane i sprawdzone przez Jönsson, która wykazała występowanie u badanych preferencji do wykorzystywania znanych, nieopartych na okulografii metod sterowania polem widzenia kamery (zob. [JJ05]). Pomimo zaobserwowanego trendu, autorzy wymienionych publikacji podkreślają pozytywny wpływ sterowania wzrokowego na poczucie zanurzenia w wirtualnej rzeczywistości u badanych osób.

Naturalną pochodną opisanego wyżej podejścia jest celowanie oparte na danych o punkcie skupienia wzroku w grach polegających na strzelaniu do przeciwników. Ta koncepcja była wielokrotnie poruszana w literaturze (por. [JJ05, IM06, LM04, SAMHV09]). Autorzy wszystkich wymienionych publikacji zwracają jednak uwagę na problem niedostatecznej precyzji systemu celowania opartego o utożsamianie kierunku widzenia, z kierunkiem prowadzenia ostrzału. Śledzenie małych celów oraz tych, które poruszają się z dużą prędkością jest wciąż kłopotliwe z racji małej dokładności współczesnych systemów okulograficznych oraz fizjologii aparatu widzenia, który wymaga wizualnego bodźca by zainicjować fazę płynnego pościgu, co czyni pozbawione przeskoków wyprzedzanie za pomocą wzroku niemalże niemożliwym [BM83].

Leyba i Malcolm przeprowadzili syntetyczne badanie celowania opartego na kierunku spojrzenia. Autorzy napotkali istotne problemy związane z niedokładnością kalibracji, którą skorygowali już po zebraniu danych. Niezależnie od korekcji, ogólna dokładność tak rozwiązanego celowania, mierzona jako liczba trafień w pożądane obiekty, była jednak istotnie niższa niż przy celowaniu z użyciem myszy, a więc rozwiązania popularnie stosowanego w tego rodzaju środowiskach [LM04]. Zarówno u Jönsson, jak i w pracy Isokoski i Martin, porównanie dokonywane jest w rzeczy-

Rysunek 2.30: Autorska gra komputerowa *Invasion*, która służy jako podstawa do przeprowadzanych badań.



wistym środowisku gry komputerowej. W obu przypadkach nie stwierdzono pozytywnego wpływu metody celowania opartej o okulograf na celność rozumianą jako liczbę skutecznych trafień [IM06, JJ05]. Jednak Jönsson zauważyła, że badane przez nią osoby uznały kontrolę z użyciem wzroku za bardziej atrakcyjną i szybszą. W obu tych przypadkach badania dotyczyły prostego dostosowania kierunku strzału do kierunku spojrzenia.

Autorzy pracy zespołu San Agustina przeprowadzili szczegółowe porównanie wydajności kanału interakcji wspartego danymi okulograficznymi z innymi metodami, w tym z, nazywaną dalej klasyczną, metodą opartą o mysz komputerową. Wykazali oni przewagę tego pierwszego w zakresie szybkości reakcji oraz zwrócili uwagę na kwestię rozmiaru aktywnych obszarów w skutecznie działających systemach interakcji. Proponują użycie odpowiednio większych obszarów, wykraczających poza sylwetkę celu, aby zredukować negatywny wpływ zaszumienia sygnału i błędów systematycznego pochodzącego z niedokładności sprzętowych i kalibracyjnych [SAMHV09].

#### 2.4.3.2 Proponowane rozwiązania

Przyjmując prostą grę komputerową za środowisko badawcze, wprowadzono do niej trzy elementy interakcji wykorzystujące dane pozyskane z pomocą okulografii:

- 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 spojrzeniem 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

W tym rozdziale zostaną poruszone przede wszystkim dwa pierwsze zagadnienia. Natomiast interaktywna symulacja głębi ostrości zostanie obszernie omówiona w Rozdziale 3.1. Przytoczone tu proponowane elementy są autorskimi rozwiązaniami opublikowanymi w artykule [BM14b].

## Środowisko testowe

Jako podstawę do badań wykorzystano autorską, zręcznościową grę komputerową pt. *Invasion* (zob. Rys. 2.30), której akcja rozgrywa się w trójwymiarowym przestrzeni. Gracz ma za zadanie odeprzeć atak wrogich statków kosmicznych korzystając z dwóch rodzajów broni. Głównym wyzwaniem jest dokładne celowanie, mając pełną swobodę rotacji w dwu płaszczyznach, a także przeciwdziałanie silnemu odrzutowi broni. Wirtualna kamera sprzężona jest z torem wystrzeliwanych pocisków. Rozgrywka trwa tak długo, dopóki gracz nie wyeliminuje wszystkich fal nacierających wrogów (wygrana) lub dopóki nie zostaną zniszczone wszystkie budynki w bazie gracza (przegrana). Za trafienia i strącenia wrogów naliczane są punkty, których suma obrazuje rezultat danej sesji gry.

Program został napisany w języku *C#* na platformę *Microsoft .NET*. Jako *API* graficzne wykorzystuje *OpenGL* wraz z możliwościami programowalnego potoku renderowania. W grze użyta została symulacja zjawiska głębi ostrości, opisana w Rozdziale 3.1.

## Selektywna prezentacja informacji kontekstowej

Charakter rozgrywki w grze *Invasion* wymaga tego, by graczowi stale był prezentowany zbiór istotnych dla niego informacji o każdym z obiektów widocznych na ekranie. Poza nazwą danego obiektu, do danych tych należy stan zniszczenia (poziom tzw. *punktów życia*), odległość od gracza oraz pułap na jakim znajduje się dany obiekt. Jednoczesne ukazanie wymienionych informacji dla wszystkich znajdujących się w danej chwili na ekranie obiektów skutkowałoby zagęszczeniem informacji, wykraczającym poza zdolności percepcyjne gracza i rozpraszącym jego uwagę (zob. Rys. 2.31a) [MB98].

Mając dostęp do wiedzy o obiekcie, na którym w danej chwili spoczywa uwaga gracza, jest możliwe dopasowanie podzbioru wyświetlanych informacji tak, by te dotyczyły wyłącznie tego obserwowanego obiektu. W proponowanym rozwiązaniu pomija się całkowicie wszystkie pozostałe elementy świata, tym samym redukując liczbę dystraktorów (zob. Rys. 2.31b i 2.31c). W ten sposób otrzymuje się też podstawowe sprzężenie zwrotne opisywanego systemu, dając użytkownikowi poczucie związania jego spojrzenia z reagującym na nie światem rozgrywki. Proponowane rozwiązanie jest implementacją paradygmatu wyświetlania zależnego od spojrzenia (ang. *gaze-contingent display*) [DC07], gdzie treść prezentowanego obrazu zależy od tego, jak jest on oglądany.

## Ruch kamery

Z uwagi na znane z przytoczonych wcześniej pozycji literaturowych kwestie związane z problematycznością prostych metod realizacji celowania opartego o spojrze-



(a)



(b)



(c)

Rysunek 2.31: Wyświetlanie wizualnej informacji kontekstowej o elementach świata gry. Nadmiar informacji (a), informacje ograniczone do obiektu na który spogląda użytkownik (b), obraz widziany przez gracza (c). Zaciemniony obszar w (a) i (b) został naniesiony w celu zwizualizowania ograniczeń widzenia peryferyjnego, nie jest on widoczny w grze.

nie, w opisywanym rozwiązaniu wykorzystano koncepcję zbliżoną do systemu *MAGIC* [ZMI99]. Zgodnie z tą ideą, dane okulograficzne używane są jedynie w celu przyspieszenia i uproszczenia interakcji, prowadząc do stanu, który można określić jako zbliżony do oczekiwanego przez użytkownika. Ten zaś ma możliwość łatwiejszego wykorzystania dotychczas znanej metody interakcji w celu precyzyjnej kontynuacji swoich działań. W opracowanej grze kamera podąża więc za wybranym przeciwnikiem, jednak gracz wciąż ma możliwość skorygowania kierunku wystrzału oraz dokładnego prowadzenia celownika z niezbędnym wyprzedzeniem za pomocą myszy. Precyzyjna interakcja dokonywana jest więc wciąż z użyciem kontrolera dokładniejszego i nieobarczonego istotnym błędem kalibracji czy zaszumienia, który to błąd jest typowy dla współczesnych systemów okulograficznych. Zaniechanie automatycznego śledzenia przez kamerę następuje, gdy wprowadzona z użyciem myszy korekcja przekroczy zadany próg, a więc gdy gracz dokona obrotu kamery tak by celownik znalazł się daleko od środka sylwetki przeciwnika. Progowa odległość została ustalona eksperymentalnie dla zapewnienia wygody i intuicyjności rozgrywki, przy czym może być zmieniana z poziomu programu.

Ruch kamery rozpoczyna się z opóźnieniem od momentu wykrycia przeniesienia uwagi wizualnej na nowy obiekt. Opóźnienie zostało wprowadzone w celu przeciwdziałania niechcianej zmianie kierunku obserwacji w momencie, gdy intencją gracza nie jest śledzenie konkretnego obiektu, lecz wykonuje on zadanie poszukiwania wizualnego. Podczas tego poszukiwania (ang. *visual search task*, por. Rozdz. 1.2) można oczekiwać, że czasy fiksacji na kolejnych obiektach wyniosą od 200 do 500 ms (por. [HE98]), stąd długość opóźnienia mającego za zadanie pominąć te fiksacje została ustalona na 500 ms.

### 2.4.3.3 Ewaluacja proponowanych rozwiązań

Celem badania było potwierdzenie hipotezy o pozytywnym wpływie proponowanych, opartych na danych okulograficznych, 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.

Ta z wersji, z którą badane osoby miały kontakt jako pierwszą, była oparta na klasycznym sterowaniu, odbywającym się wyłącznie z użyciem klawiatury i myszy. W tym przypadku to obiekt który znajdował się najbliżej środka ekranu był tym, dla którego wskazane zostały informacje i jednocześnie tym, który znajdował się w zakresie głębi ostrości. Gra pozbawiona elementów interakcji z użyciem wzroku nie oferowała także śledzenia wyróżnionego przeciwnika.

Po ukończeniu jednej sesji rozgrywki w oparciu o klasyczny schemat sterowania, badanego proszono o przebycie analogicznej sesji z wykorzystaniem rozszerzeń interfejsu opartych o dane okulograficzne. Po zakończeniu drugiej sesji, każdy z badanych otrzymywał do wypełnienia ankietę, w ramach której udzielał odpowiedzi na 5 pytań dotyczących jego subiektywnych odczuć. Odpowiedzi udzielane były za pomocą dyskretnej, 5-stopniowej skali Likerta. Każde z pytań ankiety występowało dwukrotnie, zarówno dla wersji gry pozbawionej, jak i wykorzystującej rozszerzenia oparte o dane okulograficzne.

Pytania zawarte w ankiecie i opis skali odpowiedzi dla każdego z nich:

- $Q_1$ : Jak oceniasz łatwość celowania? (*bardzo trudne – bardzo łatwe*)
- $Q_2$ : Jak oceniasz łatwość uzyskania informacji o przeciwnikach? (*bardzo trudne – bardzo łatwe*)
- $Q_3$ : Jak oceniasz intuicyjność rozgrywki? (*bardzo nieintuicyjna – bardzo intuicyjna*)
- $Q_4$ : Jak oceniasz swoje zanurzenie w świecie gry? (*bardzo małe – bardzo duże*)
- $Q_5$ : Jak oceniasz atrakcyjność rozgrywki? (*bardzo mało atrakcyjna – bardzo atrakcyjna*)

Dodatkowo zapisywano wyniki punktowe każdego z graczy uzyskane w obu sesjach gry, przedstawiające w sposób obiektywny rezultaty osiągnięte podczas rozgrywki.

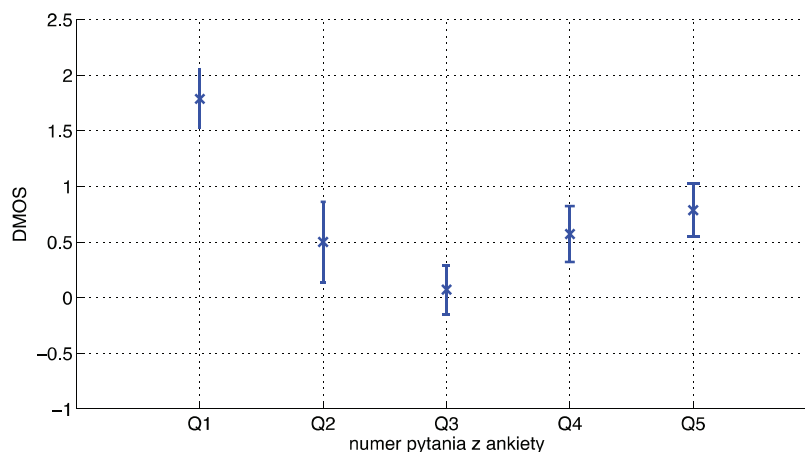
### Grupa badawcza i środowisko badań

Badanie zostało przeprowadzone na grupie 14 osób (4 kobiety i 10 mężczyzn) w wieku od 20 do 25 lat. Dodatkowo dwie osoby nie zostały uwzględnione w rezultatach, ponieważ z nieustalonych powodów nie były w stanie przejść procedury kalibracyjnej. Do właściwej rozgrywki, a tym samym udziału w eksperymencie, dopuszczone zostały te osoby które pomyślnie przeszły 5-punktową procedurę kalibracyjną. Błąd kalibracji w trakcie rozgrywki był na bieżąco oceniany przez eksperta mającego w czasie rzeczywistym podgląd sytuacji na drugim ekranie. Podczas żadnej z uwzględnianych w wynikach sesji nie zaobserwowano degradacji jakości danych mogącej mieć istotny wpływ na oceniane rezultaty.

Spośród badanych osób, cztery deklarowały zupełny brak znajomości branży gier komputerowych, cztery deklarowały jej bardzo dobrą znajomość, a pozostałe sześć umiarkowaną. Czas każdej z rozgrywek, zarówno w jednej jak i w drugiej wersji, wynosił w przybliżeniu 2 minuty, a wahania wynikały ze sposobu gry danej osoby. Całe badanie odbywało się w pomieszczeniu z kontrolowanymi warunkami oświetleniowymi, a rozgrywka poprzedzona była indywidualnym instruktażem.

Stanowisko badań (patrz Rys. 5) składało się z dwóch komputerów. Jeden z nich służył generowaniu wizualizacji wirtualnej rzeczywistości i wyposażony był w procesor *Intel i7 930* i kartę graficzną *NVIDIA GEFORCE GTX480*, drugi zaś był dedykowany przetwarzaniu danych okulograficznych. Wymiana danych pomiędzy nimi następowała w oparciu o interfejs *Ethernet*. Gracz obserwował obraz wyświetlany z częstotliwością 60 klatek na sekundę na 22-calowym monitorze *Dell E2210* o rozdzielczości 1680 na 1050 pikseli i fizycznych wymiarach aktywnego obszaru 475 na 298 milimetrów. Drugi ekran, na którym poza stanem rozgrywki wyświetlany był podgląd bieżących danych o skupieniu wzroku gracza w formie mapy cieplnej, służył operatorowi monitorującemu przebieg badania. Wykorzystany wideookulograf to urządzenie pracujące zdalnie marki *SensoMotoric Instruments*, model *RED 250*, w trakcie badania dokonujące próbkowania z częstotliwością 250 Hz. Ruch badanych osób nie był fizycznie ograniczony dla uniknięcia negatywnego wpływu na poczucie

Rysunek 2.32: Uśredniona różnica ocen (*DMOS*) wskazująca na preferencje obserwatorów w zakresie używania (wartości dodatnie) bądź nieużywania (wartości ujemne) okulografu w grze komputerowej. Kolejne słupki odpowiadają kolejnym pytaniom ankiety i obrazują odchylenie wartości *DMOS* od średniej wyrażone wartością *SEM* (ang. *Standard Error of Mean*). Ilustracja zaczerpnięta z autorskiej publikacji [BM14b].



zanurzenia i zapewnienia komfortu rozgrywki, jednak odległość oczu od ekranu była w zakresie od 50 do 70 centymetrów, co było mierzone na bieżąco przez system okulograficzny.

## Wyniki

Na potrzeby analizy statystycznej, odpowiedzi udzielane przez badanych z użyciem skali Likerta, kodowano liczbami całkowitymi z zakresu  $< 0; 4 >$ . Przy czym wartość największa oznaczała odpowiedź najbardziej pozytywną.

Zagregowane odpowiedzi zwizualizowano na Rysunku 2.32. Przedstawiono je w postaci uśrednionych różnic pomiędzy indywidualnymi odpowiedziami respondentów na pytania w obu rozpatrywanych przypadkach: gdy nie były używane, jak i gdy były używane dane okulograficzne podczas rozgrywki. Miara ta, znana jako *DMOS* (ang. *Difference Mean Opinion Score*), pozwala na pokazanie preferencji badanych osób.

Z racji niewielkiej liczebności próby (14 osób) oraz rangowego charakteru skali Likerta, do analizy wpływu opisywanych technik na odpowiedzi z ankiety posłużyły nieparametryczne metody statystyczne. Niemożność wykazania normalności rozkładów, jak i ich symetryczności względem mediany, poskutkowało koniecznością użycia metody pozbawionej takich założeń. W roli weryfikatora wykorzystano więc test znaków będący uproszczoną, pierwotną wersją testu Wilcozona dla par obserwacji. W metodzie tej dokonuje się zliczenia par zważywszy na znak różnicy wartości udzielonych odpowiedzi, a pomijając nieokreślone w skali rangowej odległości tychże różnic. W przypadku odpowiedzi na pytanie  $Q_1$ , dotyczące łatwości celowania, można zauważyć wyraźną różnicę pomiędzy odpowiedziami udzielonymi w sytuacji klasycznej kontroli i gdy użyto proponowanych w niniejszym artykule roz-



wiązań. Test znaków wykazał różnicę już na poziomie istotności  $p < 0.01$  (wartość statystyki  $Z = 3.3$ ). Można wnioskować, że automatyczne śledzenie obserwowanego obiektu przez kamerę oraz wyróżnianie przeciwnika, na którym spoczywa wzrok gracza znacząco ułatwia precyzyjne skupienie na interesującym obiekcie. Co ciekawe, ułatwienie celowania znalazło tylko nieznaczące odzwierciedlenie w osiągniętych przez graczy wynikach punktowych, których rozkład zmienił się jedynie z istotnością  $p = 0.18$  ( $Z = 1.3$ ). Wydaje się więc, że tej zmiany odczuć u badanych nie należy utożsamiać z ułatwieniem rozgrywki.

Odpowiedzi na pytanie  $Q_2$  dotyczące łatwości uzyskania informacji o przeciwnikach zdają się nie zmieniać w istotny sposób ( $p = 0.34$ ,  $Z = 0.9$ ). Oznaczałoby to, że nie można mówić o wyraźnej poprawie czytelności prezentowanej informacji kontekstowej przy dynamicznej rozgrywce, gdy powiążemy ją wyłącznie z elementem, na którym spoczywa w danej chwili uwaga wizualna obserwatora. Należy jednak pamiętać, że aspekt proponowanych rozwiązań związany z selektywną informacją kontekstową nie ma na celu wyłącznie ułatwienia pozyskiwania informacji. Jest to także sprzężenie zwrotne dla użytkownika, które ma wpływać na jego poczucie związku z prezentowanym obrazem wirtualnej rzeczywistości, co w dalszej kolejności może mieć wpływ na pogłębienie wrażenia zanurzenia w świecie gry.

Intuicyjność rozgrywki, której dotyczyło pytanie  $Q_3$ , nie wykazuje związku z opisywanymi rozwiązaniami ( $p = 0.68$ ,  $Z = 0.4$ ). Przyczyn można dopatrywać się w prostocie całokształtu zasad, nie przysparzających kłopotu także w klasycznej wersji sterowania.

Najważniejszą z punktu widzenia niniejszego artykułu jest kwestia subiektywnego odczucia zanurzenia graczy w świecie gry, której dotyczy pytanie  $Q_4$ . Na poziomie istotności  $p < 0.1$  ( $Z = 1.7$ ) udało się wykazać, że dołączenie do interfejsu graczy-gra metod opartych o wzrok, powoduje zwiększenie wspomnianego poczucia. Tłumaczyć można to tym, że środowisko rozgrywki reagujące na zachowanie gracza stwarza wrażenie większej więzi z wirtualnym otoczeniem.

Można wnioskować, że powyższe wpływa także na odpowiedzi na pytanie  $Q_5$ , a więc na wrażenie atrakcyjności rozgrywki. Jej ocena również na poziomie istotności  $p < 0.1$  ( $Z = 2.5$ ) jest wskazywana jako większa w przypadku gdy wykorzystuje się okulograf. Tutaj jednak należy dopatrywać się także wpływu swojego rodzaju nowinki, jaką jest takie wykorzystanie urządzenia do rejestracji kierunku widzenia.

Ciekawej obserwacji dokonano także poszukując z użyciem metody Spearmana korelacji pomiędzy deklarowanym przez badanych obyciem z grami komputerowymi, a poszczególnymi odpowiedziami. O ile w przypadku sterowania klasycznego ocena łatwości celowania w pytaniu  $Q_1$  jest dość silnie skorelowana z doświadczeniem gracza (współczynnik korelacji  $R = 0.6$ ), o tyle przy użyciu danych okulograficznych do jego wspomaganie korelacja ta praktycznie zanika ( $R = 0.1$ ). Można na tej podstawie wnioskować, że takie wspomaganie rozmywa zależność poczucia przystępności rozgrywki od poziomu zaznajomienia gracza z tą formą rozrywki. Istotne korelacje z doświadczeniem znaleziono także dla pytań  $Q_2$  ( $R = 0.6$ ) i  $Q_3$  ( $R = 0.7$ ).

## Wnioski i dalsze prace

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).

## Podsumowanie rozdziału

Współczesne systemy okulograficzne pozwalają na wygodne i często nieinwazyjne określenie kierunku widzenia obserwatora oglądającego wyświetlany na ekranie obraz. Częstotliwość próbkowania i niewielkie opóźnienia pozwalają na rozpoznanie fiksacji i etapów płynnego pościgu w czasie zbliżonym do rzeczywistego. Ograniczona jakość danych może być skompensowana przez zastosowanie technik filtracji danych, które pozwalają na zmniejszenie zaszumienia, ograniczenie wpływu błędu systematycznego na uzyskiwane wyniki, a także na wskazanie konkretnego obiektu na prezentowanym obrazie, któremu obserwator poświęca w danej chwili uwagę.

Wykorzystując zaproponowane techniki możliwe jest uzyskanie dostatecznej jakości informacji, która pozwala na modyfikację wyświetlanej zawartości z uwzględnieniem przebiegu uwagi obserwatora. Można więc zastanowić się w jaki sposób wiedza o części obrazu w którą spogląda użytkownik mogłaby zostać wykorzystana w procesie jego syntezy, by ostatecznie wpłynąć na sposób postrzegania tego obrazu, w tym na postrzeganie głębi przedstawianej sceny.

# Rozdział 3

## Modelowanie widzenia głębi

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.

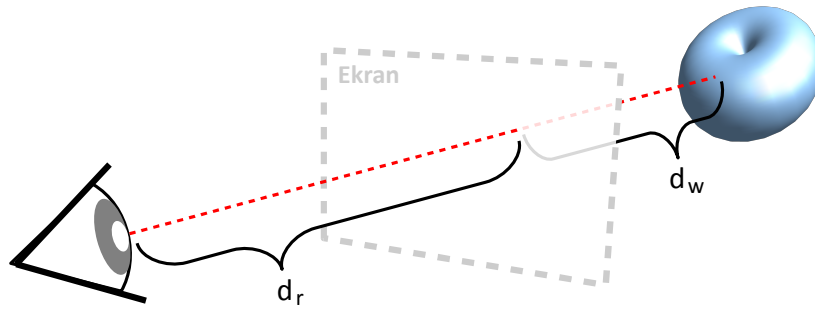
W tym rozdziale poruszone zostaną przede wszystkim kwestie związane ze zjawiskiem głębi ostrości, różnicami kontrastu i obrazowaniem stereoskopowym. Rozdział 3.1 będzie dotyczył idei interaktywnej, opartej o dane okulograficzne symulacji rozmycia wynikającego z głębi ostrości. Przedstawiona zostanie implementacja takiego rozwiązania (Rozdział 3.1.2), badanie jej skuteczności w kontekście wpływu na postrzeganie realizmu przez użytkowników (Rozdział 3.1.3) oraz łagodzenia konfliktu konwergencji i akomodacji w obrazowaniu stereoskopowym (Rozdział 3.1.4). Treści te dotyczą zagadnień będących przedmiotami współautorskich publikacji [Baz10, MBT11, Duc14].

Dalsza część Rozdziału poświęcona jest wpływowi kontrastu na postrzeganie głębi przy jednoczesnym występowaniu binokularnych wskazówek głębi na wyświetlaczach *HDR* (Rozdział 3.2.1) i jego wpływowi na wrażenie realizmu (Rozdział 3.2.2). Ta część opiera się na pracy opublikowanej w ramach współautorskiego artykułu [Pet14].

### 3.1 Interaktywna symulacja głębi ostrości

Głębina 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 (zob. Rozdział 1.1.2.2).

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



Rysunek 3.1: Dwa odrębne źródła informacji o odległości podczas oglądania obrazu wyświetlanego na ekranie: fizyczna odległość pomiędzy obserwatorem a ekranem ( $d_r$ ) oraz pozorna odległość w scenie obrazowanej na ekranie ( $d_w$ ).

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. Widzowie są przyzwyczajeni do bycia prowadzonymi przez złożone sceny w wyniku dobrania płaszczyzny ostrzenia tak, by wyostrzyć najistotniejsze elementy obrazu [Mon00]. Oddają więc kontrolę nad procesem akomodacji twórca bodźca, godząc się na utratę możliwości postrzegania ostrego obrazu na wybranej przez siebie odległości w oglądanej scenie.

Należy tu podkreślić, że oparcie się na prezentowanym w oglądanym obrazie rozmyciu mającym być efektem działania głębi ostrości, jest wyzbyciem się zdolności do intencjonalnej akomodacji wyłącznie w obrębie sceny zawartej na obrazie. Trzeba bowiem rozdzielić dwa źródła informacji o głębi: tę pochodzącą ze świata rzeczywistego (odległość obserwatora od medium, na którym obraz jest prezentowany) oraz tę ze świata przedstawionego na samym obrazie (głębina prezentowanej sceny) (por. Rys. 3.1). Oglądając obraz na płaskim medium, np. na ekranie monitora komputerowego, poprawnie funkcjonujący aparat widzenia obserwatora dokonuje akomodacji do odległości wyznaczonej przez dystans na jakim znajduje się to medium. Siedząc  $50\text{cm}$  przed ekranem widz zaakomoduje się więc do takiej właśnie odległości, niezależnie od tego jak duże dystanse są obrazowane na oglądanym przez niego bodźcu. Te ostatnie nie wywołują rzeczywistej reakcji mechanizmu akomodacji, ponieważ sam obraz znajduje się fizycznie zawsze w tej samej odległości, niezależnie od tego czy prezentuje krajobrazy z odległą linią horyzontu, czy też bardzo bliskie przedmioty zwizualizowane w dużym powiększeniu. Ten dualizm informacji o głębi niesie za sobą istotne skutki dla komfortu oglądania obrazów stereoskopowych z wymuszoną reakcją w zakresie konwergencji gałek ocznych, co zostało opisane w Rozdziale 3.1.4.

Powierzenie twórcy obrazu kontroli nad rozmyciem, wynikającym z działania głębi ostrości na odległościach w obrazowanej scenie, wiąże się z bardzo istotnym czynnikiem. Uniemożliwia ono widzowi dostrzeżenie szczegółów tych elementów sceny, które znajdują się daleko od aktualnej płaszczyzny ostrzenia, a więc są na obrazie rozmyte. Odbiorca nie ma możliwości wpłynięcia na oglądany obraz tak, by móc dostrzec interesujące go obiekty w sposób wyraźny, o ile twórca nie zaplanował wcześniej, że mają się one znaleźć w zakresie ostrości (por. Rys. 3.2). Chcąc dać obserwa-

torowi możliwość obejrzenia każdego elementu nieinteraktywnego obrazu należałoby więc naśladować obrazowanie z użyciem głębi ostrości o niemal nieskończonej długości, tym samym wyzbywając się możliwości dodania do niego tak istotnych przesłańek o trójwymiarowości prezentowanej sceny. W obu przypadkach oglądanie takiego obrazu nie odpowiada warunkom panującym w naturalnej scenerii, gdy za postrzeżoną ostrością obrazu obiektu stoi jego fizyczna odległość od oka obserwatora.



Rysunek 3.2: Przykład zastosowania rozmycia wynikającego z głębi ostrości w kinematografii. Umieszczenie tylko jednego elementu sceny w zakresie ostrości powoduje wymuszenie na widzu skupienia się właśnie na nim. Obserwator nie ma możliwości dostrzeżenia szczegółów innych elementów obrazu, np. postaci znajdujących się w centrum kadru, ponieważ są one rozmyte. Kwestię tę może rozwiązać interaktywna symulacja głębi ostrości. Źródło ilustracji: James Cameron, "Avatar", 20th Century Fox.

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. 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<sup>+</sup>17].

Kwestię interaktywnej symulacji głębi ostrości można rozbić na dwa problemy: określenie odległości płaszczyzny ostrzenia na podstawie danych o kierunku spojrzenia obserwatora oraz właściwą symulację rozmycia wynikającego z omawianego zjawiska.

### 3.1.1 Określenie płaszczyzny ostrzenia

W rozpatrywanym przypadku, gdzie obserwator ma mieć swobodną możliwość wyboru wyostrzonego obiektu poprzez skierowanie nań wzroku, konieczne jest posłużenie się zbieranymi na bieżąco danymi okulograficznymi. W najprostszym ujęciu rozwiązanie sprowadziłoby się do określenia ekranowych współrzędnych piksela, na którym skupiony jest wzrok obserwatora, a następnie odnalezienia odległości w wirtualnej scenie pomiędzy kamerą a fragmentem powierzchni obiektu, z którym dany piksel jest związany. Ta odległość stałaby się wówczas odległością płaszczyzny ostrzenia, a to co znajduje się bliżej kamery bądź dalej, ulegałoby rozmyciu z różną intensywnością.

W rzeczywistości podejście tego rodzaju obarczone jest istotnymi niedogodnościami. Przede wszystkim obszarowi uwagi wizualnej u człowieka nie da się sprowadzić do pojedynczego piksela. Pomijając już kwestię dysonansu kierunku spojrzenia i uwagi (por. Rozdz. 2.1.1), sam zakres pola widzenia obejmowany widzeniem centralnym związanym ze skupieniem wzroku jest znacznie większy. Przykładowo dla ekranu o fizycznej gęstości pikseli  $96\text{ppi}$  (pikseli na cal), przy oglądaniu obrazu z odległości  $50\text{cm}$ , kąt najostroższego widzenia człowieka  $1,5^\circ$  (ang. *foveal vision*) odpowiada obszarowi na którego średnicę w przybliżeniu składa się 50 pikseli.

Należy też zwrócić uwagę na błąd dokładności danych okulograficznych (por. Rozdz. 2.2). Niewystarczająca celność będzie często prowadzić do błędnego wyznaczenia odległości ostrzenia, gdyż estymowany punkt skupienia wzroku znajdować się będzie poza granicami sylwetki oglądanego obiektu. Będzie to tym bardziej kłopotliwe, im mniejsze ma on rozmiary na ekranie. Natomiast niewystarczająca precyzja doprowadzi z tego powodu do bardzo częstych, niejednokrotnie błędnych zmian w określaniu odległości ostrzenia. Problematyczność będzie tym bardziej znacząca, im bliżej krawędzi widocznej sylwetki obiektu spoglądać będzie obserwator. To zjawisko zostało wykorzystane przy projektowaniu, opisaną w Rozdziale 2.2.3.1, eksperymentalnej metryki jakości danych okulograficznych zbieranych na potrzeby wykorzystania w syntezie obrazu w czasie rzeczywistym.

Kolejnym kłopotem jest sytuacja, gdy obserwator intencjonalnie kieruje wzrok na zadany obiekt, jednak chce on postrzegać szczegóły dla elementów sceny, które są przezeń przysłonięte. Takie zjawisko może zachodzić dla fragmentów obrazu znajdujących się bardzo blisko obserwatora, jak na przykład zabrudzenia na szybie lub siatka płotu przez którą widz spogląda (por. Rys. 3.3). Ten kłopot nie jest możliwy do rozwiązania wprost przy prostej analizie okulograficznej, u której podstaw leży

założenie o tożsamości obiektu, na który skierowany jest wzrok, z obiektem który jest aktualnym celem uwagi obserwatora. By uporać się z tą kwestią i umożliwić użytkownikowi w pełni świadomy wybór tego czy chce oglądać na znajdujący się blisko obiekt przysłaniający tło, czy też jego celem jest obejrzenie tego tła, należałoby posłużyć się okulografią binokularną przy obrazie stereoskopowym. Na tej podstawie możliwe byłoby oszacowanie faktycznej odległości, która interesuje obserwatora (zob. Rozdz. 3.1.4.2).

Innym podejściem, które pozwala zredukować kłopotliwość sytuacji związanych z niewystarczającą dokładnością danych, jak i sytuacji z obiektami przysłaniającymi istotny dla obserwatora obraz, może być opisana w Rozdziale 2.3 metoda *GDOT*. Wówczas na podstawie danych okulograficznych nie tyle znajdowana byłaby konkretna odległość punktu skupienia wzroku, lecz znajdowany byłby interesujący obserwatora obiekt lub jego część (*atraktor uwagi*), której uśredniona odległość od wirtualnej kamery byłaby dopiero przyjmowana za odległość ostrzenia. Należy wówczas uważnie dobrać parametry głębi ostrości tak, by cały atraktor uwagi znajdował się w zakresie widzenia powiązany z obrazowaniem rozumianym jako wyraźne.

### 3.1.2 Symulacja rozmycia obrazu związanego ze zjawiskiem głębi ostrości

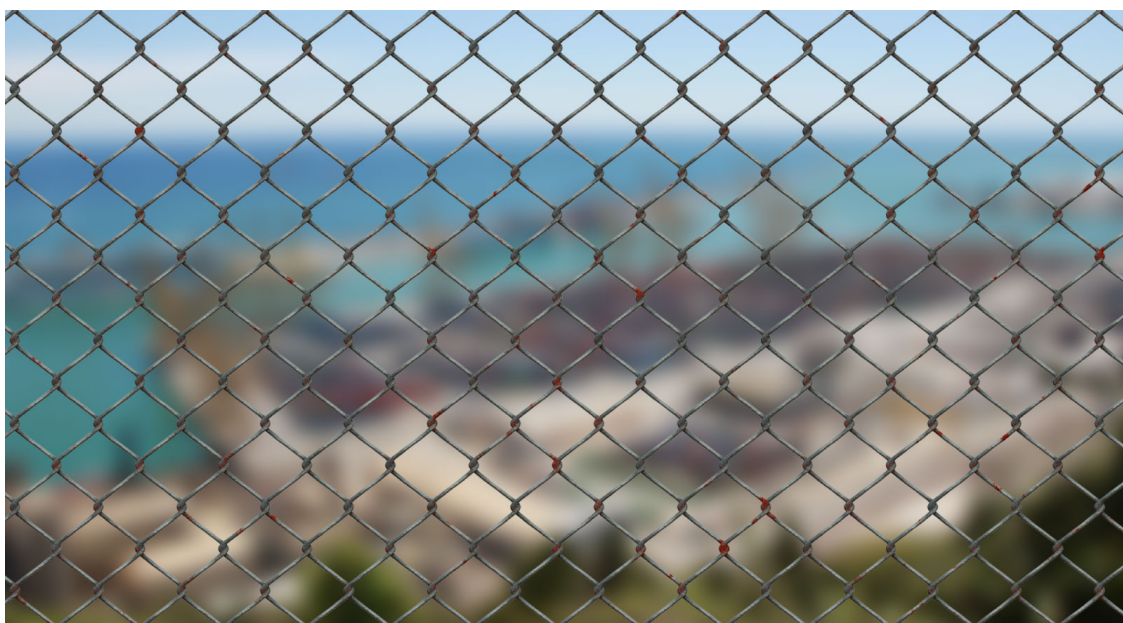
Problem symulacji rozmycia obrazu wynikającego ze zjawiska głębi ostrości jest szeroko omawiany w literaturze. Powstała duża liczba metod pozwalających z większą bądź mniejszą dokładnością naśladować wynik tego zjawiska optycznego. Barsky i Kosloff w swojej przekrojowej pracy [BK08] dokonują podziału tychże z uwagi na przestrzeń i zakres wejściowej informacji, na jakich operują. Wyróżniają metody działające w przestrzeni obiektów (ang. *Object-Space*) oraz w przestrzeni obrazu (ang. *Image-Space*). Bardziej szczegółowy podział został zaproponowany przez Demersa [Dem04], jednak wyróżnione przez niego grupy metod stanowią dokładniejszy podział systematyki Barky’ego i Kosloff’a.

#### 3.1.2.1 Metody operujące w przestrzeni obrazu

Metody działające w przestrzeni obrazu działają na gotowej klatce animacji i sprowadzają symulację do problemu rozmycia gotowej klatki. Symulacja głębi ostrości staje się więc efektem typu *post-processing* (zob. Rys. 3.4a). Takie podejście pozwala na stosunkowo łatwe dodanie efektu głębi ostrości do już istniejącego potoku renderowania, często jest też łatwiejsze do zaimplementowania w sposób wydajny korzystając z programowalnego procesora graficznego.

Demers dzieli metody bazujące na gotowej klatce na takie, które charakteryzują się podejściem rozsiewającym (ang. *scattering*) i gromadzącym (ang. *gathering*) [Dem04]. Te pierwsze prezentują ideologię obliczeń, gdzie pojedyncze cząstki danych z obrazu źródłowego wpływają na wiele elementów z docelowego zbioru, jest to więc mapowanie wprzód. Natomiast podejście gromadzące jest odwrotne: iteruje się po docelowym zbiorze, wyliczając wartości na podstawie wielu elementów z danych źródłowych, a więc mapuje się wstecznie. W ostatnich latach popularność zyskało także podejście *scatter-as-gather*, które łączy idee zawarte w obu tych





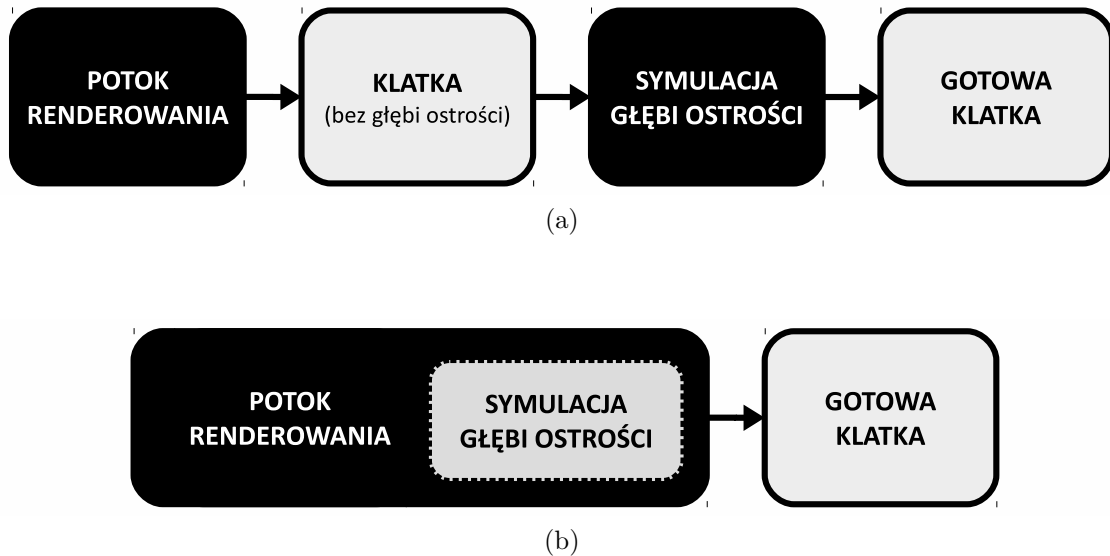
(a)



(b)

Rysunek 3.3: Problematiczny przypadek dla prostego korzystania z punktu skupienia wzroku w celu wyznaczenia odległości płaszczyzny ostrzenia. W pierwszym przypadku obserwator chce spoglądać na siatkę płotu (a), w drugim zaś na znajdujące się za nią tło (b). Oglądający może jednak celowo umieszczać punkt skupienia swojego wzroku na jej obszarze, chcąc jednocześnie obejrzeć szczegóły tła. Takie sytuacje są nie do odróżnienia w klasycznym, uproszczonym podejściu do analizy okulograficznej.





Rysunek 3.4: Dwa podejścia do integracji modułu symulującego głębię ostrości z potokiem renderowania: podejście typu *post-processing*, gdzie potok pozostaje niezmienny a symulacja operuje na gotowej klatce animacji (a) oraz podejście z głęboką integracją, gdzie symulacja wymaga zmian w samym potoku (b).

grupach [Sou13].

Rozmycie komputerowego obrazu rastrowego jest przykładem dolnoprzepustowej filtracji dwuwymiarowego sygnału. W najprostszym ujęciu, w wynikowym obrazie wartość każdego piksela jest funkcją wartości pikseli z jego otoczenia w obrazie źródłowym. Uzyskanie wartości wynikowych pikseli polega więc na wielokrotnym próbkowaniu wejściowego obrazu, a następnie wyliczeniu ważonej sumy wyników tych próbkowań. Ten schemat operacji można wygodnie zrealizować w postaci metody gromadzącej, a zatem w łatwy sposób użyć go na etapie cieniowania fragmentów (ang. *fragment shading*) w programowalnym potoku renderowania czasu rzeczywistego.

### Filtracja splotowa

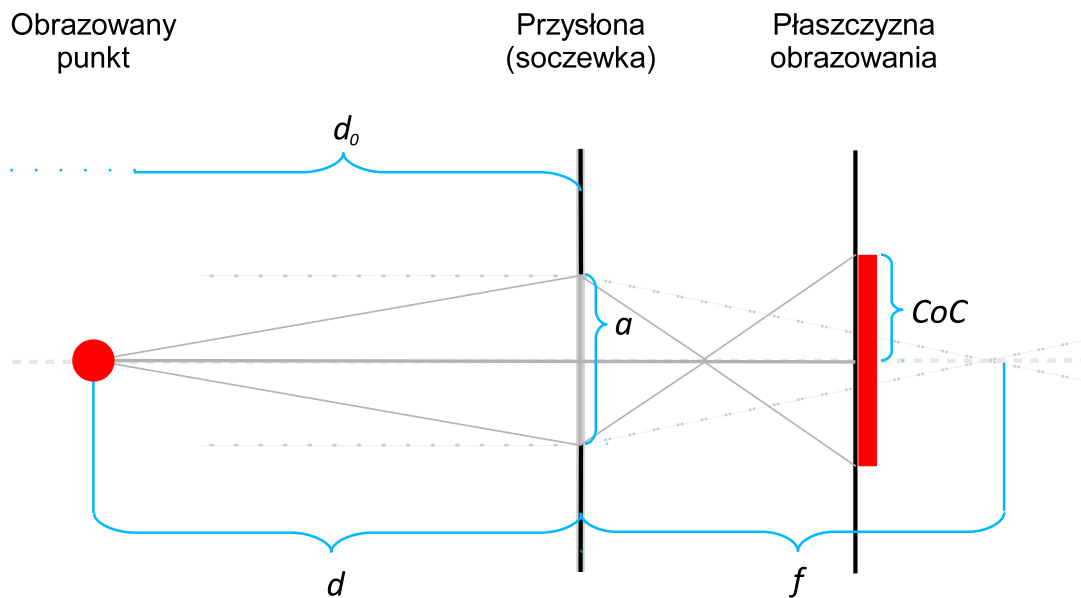
Najczęściej używanym sposobem realizacji rozmycia obrazu rastrowego jest filtracja splotowa. Obliczając wartości poszczególnych pikseli stosuje się maskę o stałych wymiarach, która determinuje zasięg próbkowania, a więc rozmiar otoczenia biorącego udział w uzyskaniu wyniku. Wagi maski zazwyczaj sumują się do wartości 1, by na wyjściu zachować wrażenie jasności wejściowego obrazu. Wartości wag mogą mieć różne pochodzenie: mogą być jednakowe dając tym samym w rezultacie uśrednienie, mogą faworyzować centralną wartość piksela [Rok96] lub np. być wartościami dwuwymiarowej funkcji Gaussa [Sze10]. Warty wyróżnienia jest również zorientowane na efektywność implementacji na współczesnych procesorach graficznych, dające nieco niższą wynikową jakość obrazu, podejście zaproponowane przez Kawase [Kaw03].

Dla syntezy obrazu w czasie rzeczywistym istotną cechą dwuwymiarowej filtracji splotowej z wagami pochodzącymi z rozkładu izotropowego, takiego jak środko-

wosymetryczna dwuwymiarowa funkcja Gaussowska, jest możliwość rozdzielania jej na dwa następujące kolejno po sobie, jednowymiarowe etapy. W ten sposób znacząco redukuje się liczbę kosztownych próbkowań oraz złożoność obliczeniową. Rozmycie z użyciem splotu realizuje się więc najpierw tylko w poziomie, by jego rezultat poddać rozmyciu pionowemu - lub na odwrót. Tym samym dla przykładowej maski o wymiarach  $5 \times 5$ , zamiast 25 próbkowań, powiązanych z nimi mnożeń i sumowań, wykonuje się zaledwie po 5 na każdym z dwóch etapów [Str14].

## Model układu optycznego

Rozmycie wynikające z głębi ostrości nie jest jednakowe dla całego obrazu będącego projekcją trójwymiarowej sceny. O intensywności rozmycia danego punktu obrazu decyduje rozmiar, powstałego przy obrazowaniu punktu, krążka rozmycia. Ten zaś zależy od różnicy pomiędzy odległością punktu od kamery, a odległością ostrzenia.



Rysunek 3.5: Model jednosoczewkowego układu optycznego.  $f$  - ogniskowa,  $d$  - odległość obrazowanego punktu od soczewki,  $d_0$  - odległość ostrzenia,  $a$  - przysłona,  $CoC$  - promień krążka rozmycia.

Zamodelowanie rozmiaru krążka rozmycia w rzeczywistym układzie optycznym jest procesem złożonym i często niepotrzebnym dla uzyskania satysfakcjonującej dla odbiorcy symulacji głębi ostrości. Najczęściej stosuje się uproszczenie polegające na posłużeniu się modelem układu o znanym rozmiarze przysłony  $a$  i z jedną soczewką o grubości bliskiej zera i znanej ogniskowej  $f$ . Układ taki znajduje się w stanie ustawienia ostrości na zadaną odległość  $d_0$  od soczewki, a rozpatrywany w danej chwili punkt sceny znajduje się w odległości  $d$  od soczewki (zob. Rys. 3.5). W ten sposób możliwe jest obliczenie promienia krążka rozmycia  $CoC$  korzystając z następującego wzoru [PC82]:

$$CoC = \left| a \cdot \frac{f}{d_0 - f} \cdot \left( 1 - \frac{d_0}{d} \right) \right|.$$

Zależność ta nie jest liniowa: zmiany rozmiaru krążka rozmycia przy małych odległościach od kamery są dużo większe, niż przy dużych odległościach (por. Rys. 3.6). Niekiedy widniejąca we wzorze wartość bezwzględna może zostać zniesiona w celu uzyskania wartości  $CoC$  ze znakiem. W ten sposób możliwe jest odróżnienie rozmycia obiektów bliżej kamery niż płaszczyzna ostrzenia, od rozmycia obiektów położonych dalej od tejże płaszczyzny. Dzięki temu można zastosować odmienne metody dla tych dwóch przypadków. Może to pomóc w redukcji artefaktów wynikających z próbkowania, gdy promień  $CoC$  jest duży, co zdarza się znacznie częściej w pobliżu kamery.

## Maska rozmycia

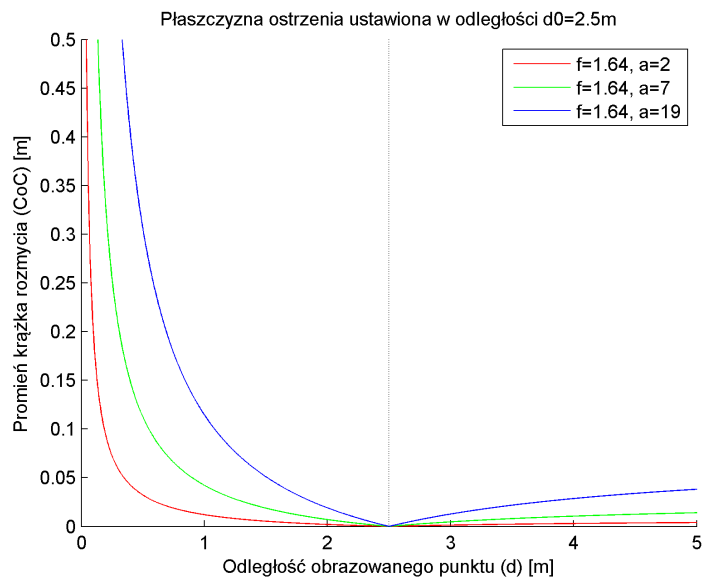
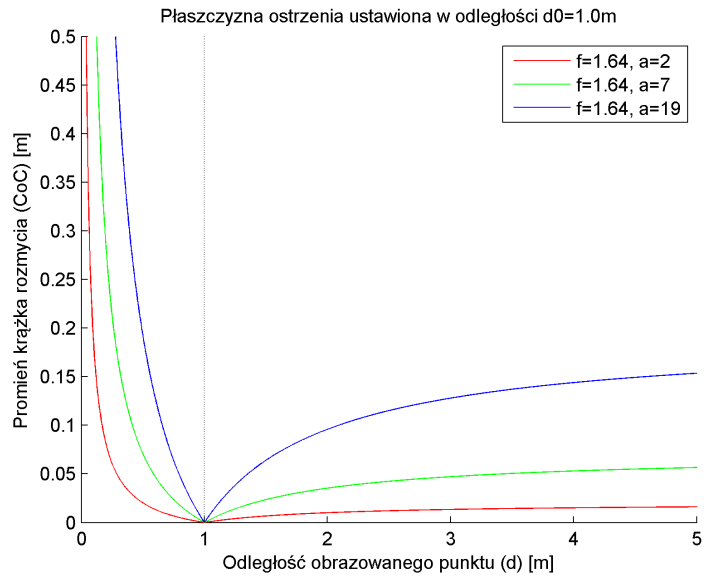
Najkorzystniejszym rozwiązaniem byłyby metoda rozmycia, w której maska, określająca miejsca próbkowania sąsiedztwa, może być dowolnie skalowana w celu pokrycia obszaru pożądanego krążka rozmycia. Współczynnikiem skalowania byłby jego promień, a więc wartość  $CoC$  obliczona dla rozpatrywanego piksela z użyciem obranego modelu układu optycznego. Należy więc mieć na uwadze, że zasięg rozmycia, co przekłada się na skalę, byłby zmienny pomiędzy pikselami obrazu.

Współczesny, sprzętowo realizowany programowalny potok renderowania czasu rzeczywistego nie pozwala na efektywne wykorzystanie zmiennej liczby próbkowań w jednym przebiegu [Kes16].

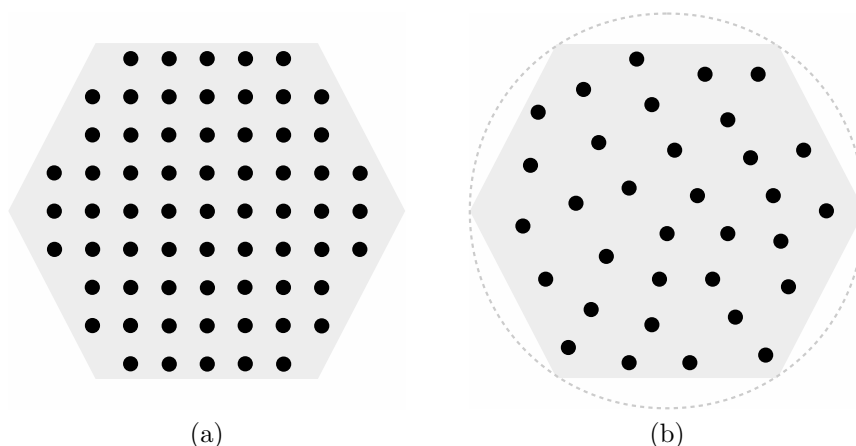
Możliwe jest przybliżenie rozmycia o żądanym zasięgu poprzez iteracyjny proces składający się z wielu etapów korzystających z tej samej maski o stałym rozmiarze. W ten sposób, wielokrotnie powtarzając rozmycie np. z użyciem niewielkiej maski o wymiarach  $3 \times 3$ , można zasymulować krążek rozmycia dowolnych rozmiarów. Liczba powtórzeń dla całego obrazu jest wówczas zależna od największej występującej w nim wartości  $CoC$  [Rok97]. Duże maski są mniej efektywne przy implementacji opartej o współczesny programowalny potok renderowania z uwagi na mniejszą możliwość wykorzystania pamięci podręcznej tekstur (ang. *texture cache*). By zredukować liczbę potrzebnych powtórzeń dla uzyskania większego zasięgu rozmycia, a więc odpowiednika większej maski, można posłużyć się uprzednio pomniejszoną wersją oryginalnej klatki [Wol13, Str14].

Omówione powyżej sposoby realizacji rozmycia bazują na założeniu o okrągłym, środkowosymetrycznym kształcie krążka rozmycia. Wzorując się na rzeczywistym mechanizmie powstawania zjawiska głębi ostrości, należy wziąć pod uwagę, że kształt odwzorowania punktu na płaszczyźnie obrazowania zależy głównie od kształtu otworu przysłony. W powszechnie wykorzystywanych aparatach fotograficznych i kamerach kształt ten zazwyczaj nie jest kolisty, a raczej jest zbliżony do foremnego wielokąta. Ma to związek z najczęstszym mechanizmem działania przysłony, która jest realizowana jako tzw. diafragma czyli zestaw  $n$  listków pozwalających na jej przyamykanie do dowolnej wielkości otworu.

Zatem w fotografii i kinematografii najczęściej spotykanym kształtem krążka rozmycia jest foremny wielokąt. Zjawisko gdzie ten kształt jest wyraźnie widoczny



Rysunek 3.6: Nieliniowość zależności między odległością obrazowanego punktu względem odległości ostrzenia, a rozmiarem krążka rozmycia w jednosoczewkowym modelu układu optycznego. Na wykresach pionową, przerywaną linią oznaczono odległość płaszczyzny ostrzenia od kamery. Kolorem czerwonym, zielonym i niebieskim ukazano promień krążka rozmycia dla obrazowanych punktów znajdujących się w różnej odległości od kamery, przy różnych wielkościach przysłony  $a$  i stałej wartości ogniskowej  $f$ . W przypadku (a) ostrość jest ustawiona na odległość  $d_0 = 1m$ , w przypadku (b)  $d_0 = 2,5m$ .



Rysunek 3.7: Próbkowanie kształtu krążka rozmycia. Regularna siatka (a), dysk Poissona (b).

na zdjęciu, nazywa się popularnie *bokeh* (z języka japońskiego). Dlatego chcąc naśladować fotograficzne lub filmowe rozmycie wynikające z głębi ostrości, należy wziąć pod uwagę symulowanie takiego właśnie kształtu.

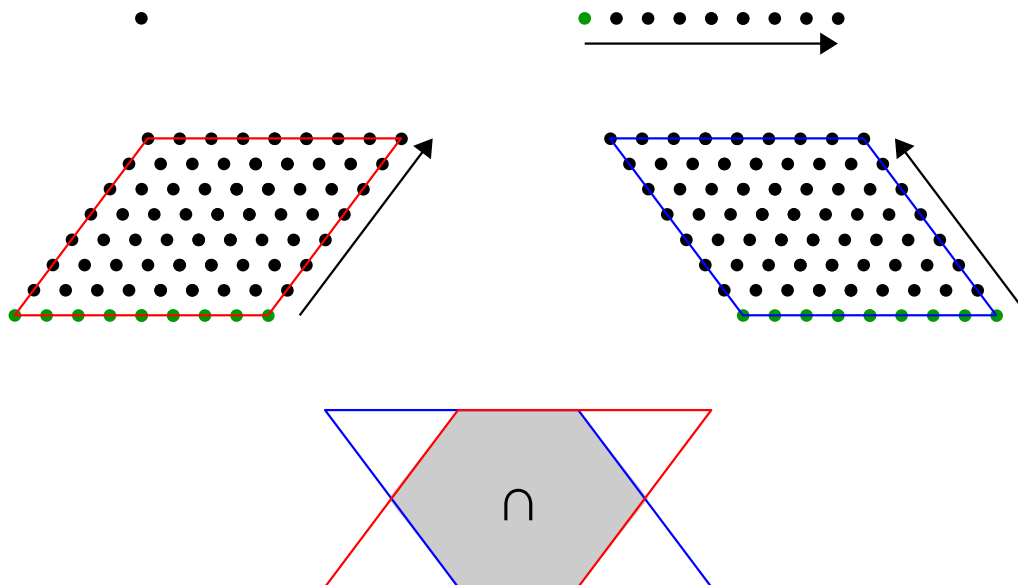
Stosując podejście gromadzące (*gathering*), możliwe jest przybliżenie dowolnego kształtu poprzez zastosowanie maski zawierającej próbki jedynie w obszarze tego kształtu (por. Rys. 3.7a). Przy regularnej siatce próbek takie podejście wymaga jednak wielu odczytów z tekstury dla każdego piksela wynikowego obrazu i przez to nie jest efektywne (por. [MRD12], *Naive Approach*).

Modyfikacją tej metody może być użycie nieregularnej maski powstałej w wyniku próbkowania dysku Poissona (ang. *Poisson disc sampling*) [ST04]. W wyniku takiego próbkowania otrzymuje się zbiór punktów o zadanej liczności, którego elementy nie są oddalone od siebie o mniej niż zadany próg. Dzięki temu możliwe jest uzyskanie z pomocą ograniczonej liczby punktów pokrycia obszaru o zadanym kształcie w sposób przybliżony do jednorodnego, przy czym rozkład ten jest bardziej naturalny od prostej siatki (por. Rys. 3.7b) [Bri07]. Z uwagi na kosztowność obliczenia położenia próbek w takim podejściu, często dokonuje się tego przed rozpoczęciem właściwego renderowania i nie zmienia się jej w trakcie generowania klatek animacji.

W literaturze można również znaleźć podejścia wykorzystujące rozdzielanie próbkowań w celu pokrycia zadanego kształtu, które jest ideowo zbliżone do rozdzielania filtracji Gaussowskiej na oddzielne przebiegi poziome i pionowe. Z tą różnicą, że docelowy kształt powstaje w wyniku więcej niż dwóch etapów, a kolejne przebiegi dotyczą próbkowania pod różnym kątem, niekoniecznie poziomym i pionowym (por. Rys. 3.8) [MRD12, WBB11].

Występują też podejścia w których upraszcza się sytuację i rozmywa się obraz z użyciem kilku wcześniej zdefiniowanych wartości *CoC* otrzymując tym samym stały rozmiar maski. Obliczenie właściwego obrazu wynikowego następuje wówczas poprzez interpolowanie wartości odczytanych z poszczególnych stopni rozmycia, zależnie od faktycznie obliczonej wartości *CoC* dla danego piksela [EH08].

Kolejną techniką jest podejście, w którym na etapie *post-processingu* budowana jest dodatkowa geometria. Dla każdego piksela źródłowego obrazu renderuje się fi-



Rysunek 3.8: Kolejne etapy rozdzielanego podejścia do próbkowania kształtu krążka rozmycia [MRD12, WBB11] (kolejno od góry, od lewej): pojedyncza próbka źródłowa, pierwszy przebieg rozmycia, przebiegi rozmycia pod dwoma różnymi kątami, kompozycja heksagonalnego kształtu jako części wspólnej uzyskanych masek.

gurę (najczęściej czworokąt) o rozmiarze zależnym od obliczonego promienia  $CoC$ . Technika ta może wykorzystywać też tzw. *point sprites*, natomiast po rasteryzacji na tej dodatkowej geometrii mapowana jest tekstura pozwalająca na uzyskanie zmiękzonego brzegu (np. wartości z funkcji Gaussowskiej) lub pożądanego kształtu *bokeh* [MD11]. Modulacja koloru zależy od wartości źródłowego piksela. Jest to podejście należące do grupy rozsiwających (*scattering*), kosztowne z uwagi na licznosc dodatkowych elementów geometrycznych, dlatego istnieją jego odmiany gdzie taki zabieg stosuje się wyłącznie od zadanej progowej wartości  $CoC$  [Fut12].

## Artefakty

Uproszczone metody symulacji głębi ostrości często skutkują powstaniem charakterystycznych niedoskonałości wizualnych. Większość z nich powstaje w rejonach nieciągłości głębi (ang. *depth discontinuities*), a więc przy granicach obiektów, które przesłaniają inne znajdujące się na trójwymiarowej scenie. W metodach gromadzących, rozmywając obraz próbkuje się otoczenie docelowego punktu. Mając do dyspozycji jedynie dwuwymiarową projekcję sceny jako zbiór danych wejściowych, otoczenie w obszarach nieciągłości głębi zawiera piksele, których wartości zostały wyliczone na podstawie punktów pochodzących z innych obiektów niż ten, który jest rozpatrywany. Może to więc prowadzić do rozmycia także tych części obrazu, które powinny znajdować się w obszarze ostrości.

Ten problem może wystąpić w dwu przypadkach. Pierwszy z nich dotyczy sytuacji, gdy przy rozmywaniu brane są pod uwagę niepożądane piksele związane z obiektami znajdującymi się bliżej od faktycznie rozmytego obiektu. Jest to więc przypadek, gdy wyraźne obiekty przesłaniają rozmyte tło. Wówczas stosując naiwną metodę, krawędzie bliższego, wyraźnego obiektu uległyby niepotrzebnemu rozmyciu



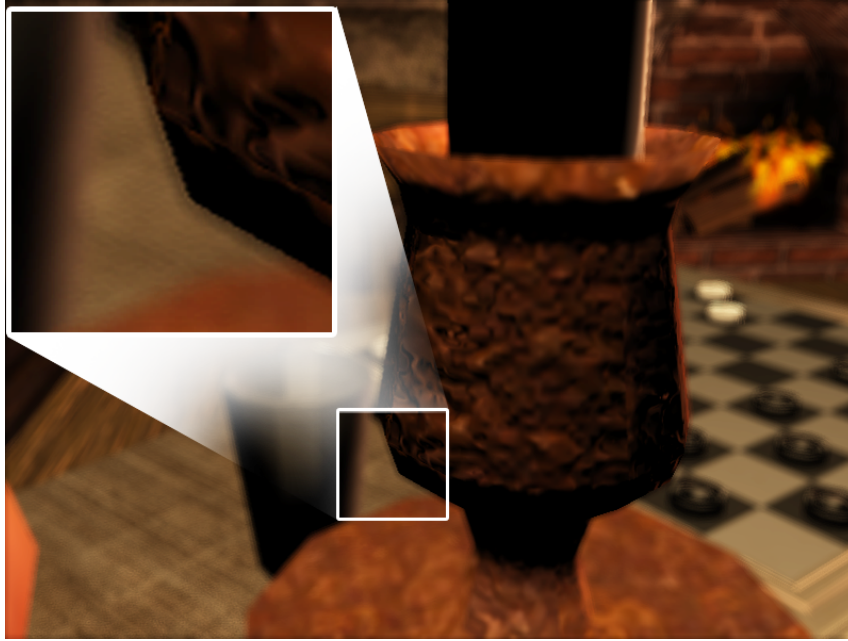
(a)



(b)

Rysunek 3.9: 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.





Rysunek 3.10: Zjawisko wycieku ostrości (ang. *intensity leakage*) powstałe przy użyciu gromadzącej (ang. *gathering*) metody rozmycia, gdy wyraźny obiekt pierwszoplanowy przesłania nieostre tło.

(zob. Rys. 3.10). Zjawisko takie nazywa się często wyciekaniem ostrości (ang. *intensity leakage* lub *intensity leaking*).

By zredukować problem wycieku ostrości, podczas próbkowania otoczenia pikselela w procesie rozmywania, należy dokonywać każdorazowo testu głębokości w celu pominięcia próbek, gdy ich odległość od kamery wskazuje na powstanie tego problemu.

Sytuacja przeciwna występuje, gdy bliższy, pierwszoplanowy obiekt znajduje się bliżej kamery niż płaszczyzna ostrzenia, a zatem powinien zostać rozmyty, natomiast jego tło jest wyraźne. Przy stosowaniu metody gromadzącej, obliczając znajdujące się w pobliżu granicy rozmytego obiektu pierwszoplanowego piksele tła okaże się, że wielkość  $CoC$  odpowiada wyraźnemu obszarowi. W efekcie granica nieostrego obiektu nie zostanie rozmyta, a zatem nie wystąpi oczekiwane zjawisko półprzezroczystej krawędzi (zob. Rys. 3.11a).

Dla metody gromadzącej próbą redukcji problemu wizualnego może być albo dodatkowy przebieg rozmycia [Dem04], albo podejście poszerzające zasięg wartości  $CoC$  dla obiektów bliżej kamery od płaszczyzny ostrości. W tym drugim przypadku konieczne staje się rozróżnienie między wartościami  $CoC$  dla bliży i dali, na przykład poprzez obliczenie ich ze znakiem. Następnie wykonuje się dodatkowy przebieg *post processing*, gdzie dokonuje się ekspansji bliskich wartości  $CoC_b$  [EH08]:

$$CoC_b^e = 2 \cdot \max(CoC_b, CoC_b^r) - CoC_b.$$

Po zastosowaniu takiego zabiegu  $CoC_b^e$  przyjmie wartości większe od zera także dla obszarów wyraźnych, jeśli w ich bliskim sąsiedztwie znajdują się, będące bliżej, rozmyte obiekty. Tym samym krawędzi nieostrych obiektów pierwszoplanowych zo-



staną rozmyte nawet wtedy, gdy znajdują się na wyraźnym tle (zob. Rys. 3.11b). Metoda ta jest jednak obciążona istotną wadą. W rozmyciu wokół obszaru oryginalnie zajmowanego przez nieostry obiekt, biorą udział także piksele pochodzące z wyraźnego tła (zob. Rys. 3.11c). Obejście tego problemu wymagałoby wprowadzenia dodatkowego etapu w procesie renderowania i komponowania finalnego obrazu poprzez nałożenie na klatkę zawierającą wyłącznie tło, części klatki z rozmytymi obiektami pierwszoplanowymi.

Warto zaznaczyć, że metody rozsiewające są w dużo mniejszym stopniu podatne na powstawanie tego rodzaju artefaktów, ponieważ sposób ich działania lepiej odpowiada rzeczywistej genezie zjawiska: rozmycie następuje w wyniku nakładania się na siebie obszarowych kontrybucji źródłowych punktów. Należy jednak pamiętać, że i w takiej sytuacji w wejściowej klatce brakuje informacji o obiektach przesłoniętych, więc rozmycie tła zasłoniętego przez wyraźny obiekt nigdy nie będzie poprawne przy podejściu *post processing* bez integracji z potokiem syntezy obrazu. Należy również pamiętać o teście głębokości podczas rozsiewania wartości, by rozmycie tła nie pojawiło się przed bliższymi obiektami.

### 3.1.2.2 Metody operujące w przestrzeni obiektów

Jest to zbiór metod, które wymagają głębokiej integracji z systemem syntezy obrazu (zob. Rys. 3.4b). Do swojego działania potrzebują informacji o scenie, a nie jedynie już wyrenderowanej klatki animacji. Metody te często pozwalają na uniknięcie problemów charakterystycznych dla metod działających w przestrzeni obrazu (zob. Rozdz. 3.1.2.1), jednak są znacznie bardziej kosztowne obliczeniowo.

### Śledzenie promieni

Śledzenie promieni jest bardzo uniwersalną metodą obliczeniową pozwalającą na symulację zachowania oświetlenia, a co za tym idzie pozyskiwania obrazu, w trójwymiarowych scenach. Polega na śledzeniu w przestrzeni sceny przebiegu promieni, wysyłanych po jednym dla każdego piksela bufora klatki. Gdy zamiast pojedynczego promienia, dla każdego z pikseli prześledzić ich więcej, pokrywając tym samym obszar symulowanej przysłony, uśrednione ich wartości pozwoliłyby uzyskać zjawisko naśladujące faktyczne powstawanie zjawiska głębi ostrości.

Metoda ta jest jednak wymagająca obliczeniowo i trudna do zaimplementowania na współczesnym sprzęcie przeznaczonym do syntezy grafiki czasu rzeczywistego. Dlatego nie wykorzystuje się jej w tym obszarze, jest natomiast popularna w grafice realistycznej.

Symulacja głębi ostrości z użyciem śledzenia promieni może bazować albo na prostym modelu układu jednosoczewkowego [CPC84] (jak w Rozdz. 3.1.2.1), albo na bardziej złożonym modelu. Pozwala więc ona na odwzorowanie optycznych właściwości rzeczywistego obiektywu [KMH95].

### Akumulacja klatek

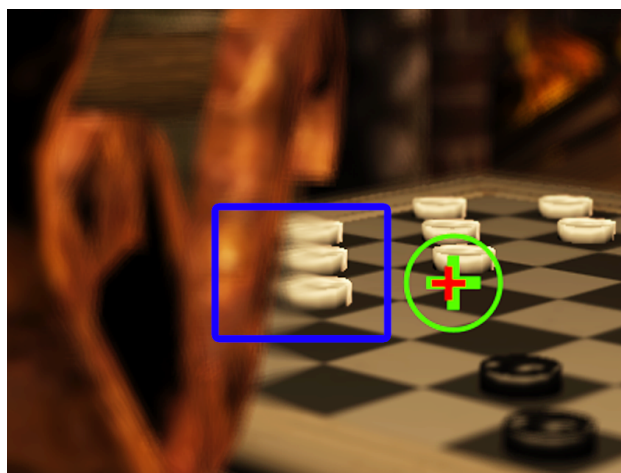
Stosunkowo dokładną metodą, lecz dostosowaną do współczesnych warunków w środowisku systemów grafiki czasu rzeczywistego, jest podejście oparte na wy-



(a)



(b)



(c)

Rysunek 3.11: Zjawisko nieciągłości głębi (a), ograniczenie jego występowania w wyniku modyfikacji algorytmu [EH08] (b) i przykład spowodowanego przez tę modyfikację wizualnego artefaktu polegającego na rozmyciu tła które powinno być wyraźne (c)

korzystaniu *bufora akumulacji* (ang. *accumulation buffer*), nazwane przez Demersa w jego systematyce podejściem wielokamerowym (ang. *multiple-camera*) [Dem04]. Podejście to wymaga wielokrotnego wyrenderowania sceny, przemieszczając wirtualną kamerę o niewielkie odległości tak, by dokonać próbkowania obszaru zbliżonego do rozmiaru symulowanej przysłony. Tak uzyskane, pozbawione rozmycia obrazy komponowane są ze sobą [HA90]. Metoda ta nie generuje artefaktów związanych z nieciągłością głębi, gdyż rozmycie powstaje wskutek efektu paralaksy wynikającego z rzeczywistego przemieszczania się obserwatora. Tym samym przysłonięcie przez obiekt bliższy kamery nie powoduje utraty wiedzy o tle.

Metoda wykorzystująca akumulację klatek, choć dająca bardzo dobrej jakości rezultaty, jest mało efektywna obliczeniowo. Wielokrotne, często kilkudziesięciokrotne, powtarzanie renderowania każdej klatki prowadzi do znaczącego wydłużenia czasu przygotowania kolejnych obrazów. Tym samym nie jest to podejście współcześnie wykorzystywane w grafice czasu rzeczywistego do innych zastosowań, niż jako referencja przy porównywaniu innych, szybszych metod.

## Renderowanie warstwowe

Zbiór metod związanych z warstwami, choć przez Barsky'ego i Kosloffa zaliczony do metod działających w przestrzeni obrazu, w niektórych przypadkach wymaga głębokiej integracji z potokiem renderowania. Wymaga bowiem utworzenia kilku buforów, w których zapisywane będą wartości pikseli powstającej klatki. Decyzja o tym, do którego z nich ma trafić dana wartość, podejmowana jest na podstawie odległości od kamery i może następować albo po dokonaniu testu głębokości, albo jeszcze przed nim. W tym ostatnim przypadku umożliwia to późniejszy dostęp do wiedzy o elementach które z punktu widzenia kamery są przysłonięte, jednak kosztem zwiększonej ilości obliczeń (np. cieniowanie przysłoniętych fragmentów, późniejszy dodatkowy test głębokości przy kompozycji ostatecznej klatki z rozmyciem).

Najprostsze podejście, dalekie od naśladowania rzeczywistego zjawiska optycznego, polega na wprowadzeniu trzech planów o indywidualnych i stałych stopniach rozmycia: bliskiego nieostrego, wyraźnego i dalekiego nieostrego. Utworzone warstwy mogą też wynikać z jednorodnej kwantyzacji zakresu prezentowanych w scenie odległości [Sco92, BTHC03]. Mogą również zależeć od wcześniej wyliczonej zmienności rozmiarów krążka rozmycia w zadanej scenie [LES09]. W literaturze można znaleźć też podejście, gdzie renderowanie do poszczególnych warstw odbywa się na zasadzie *depth peeling* [SW12].

### 3.1.2.3 Implementacja

Dokonano implementacji systemu wirtualnej rzeczywistości, który symuluje zjawisko głębi ostrości w sposób interaktywny. Celem było naśladowanie zachowania ludzkiego systemu akomodacji i zbadanie preferencji użytkowników odnośnie koncepcji interaktywności opartej o zbierane na bieżąco dane okulograficzne. Szczegóły implementacji zostały opisane na łamach dwóch autorskich publikacji [Baz10, MBT11].

Całość została zrealizowana w formie aplikacji *Microsoft .NET 3.5* napisanej w języku *C#*, wykorzystującej programowalny potok renderowania za pośrednictwem *OpenGL 3.2*. Jako źródło danych okulograficznych wykorzystano eye tracker

*SensoMotoric Instruments RED250* pozwalający na zbieranie próbek z częstotliwością 250Hz. Podczas działania program wykorzystuje dwa ekrany: ekran operatora pozwalający na dobór parametrów symulacji i podgląd danych otrzymywanych z okulografu oraz ekran użytkownika, gdzie wyświetlana jest wizualizacja wirtualnego środowiska zawierająca symulację głębi ostrości.

Moduł odbierający dane z systemu okulograficznego został zrealizowany w oparciu o protokół *TCP/IP*. Komunikacja odbywała się w sposób asynchroniczny, z poziomu osobnego wątku działającego niezależnie od procesu renderowania. Dane okulograficzne filtrowane były z użyciem metody *I-DT* (zob. Rozdz. 2.2.4.1).

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 równych wagach. Dla rozmycia każdego piksela wynikowego obrazu, gromadzono okoliczne wartości z użyciem maski o 36 współrzędnych próbkowania, co było wystarczające dla uzyskania akceptowalnej jakości rozmycia. Z uwagi na jakość, ustalono wartość maksymalną  $CoC = 12$ , by ograniczyć rozmiar  $CoC$  dla ekstremalnych sytuacji.

Implementację oparto o programowalny potok renderowania. Kolejne etapy symulacji głębi ostrości realizowano jako programy cieniujące (ang. *shading programs*). Wyniki obliczeń  $CoC$  przechowywano w formie tekstury w pamięci karty graficznej (zob. Rys. 3.12). Dzięki wykorzystaniu programowalnego procesora graficznego, renderowanie kolejnych klatek odbywało się ze stabilną częstotliwością zbliżoną do 60Hz.

Bufor przechowujący wartości  $CoC$  związane z obiektami znajdującymi się bliżej kamery niż płaszczyzna ostrości miał wymiary 8-krotnie zmniejszone w każdym z wymiarów względem oryginalnej klatki. Dzięki temu możliwe było uzyskanie lepszej jakości rozmycia przy użyciu tej samej liczby próbek, dla większego zasięgu maski dzięki wykorzystaniu sprzętowej filtracji tekstury metodą dwuliniową (ang. *bilinear filtering*).

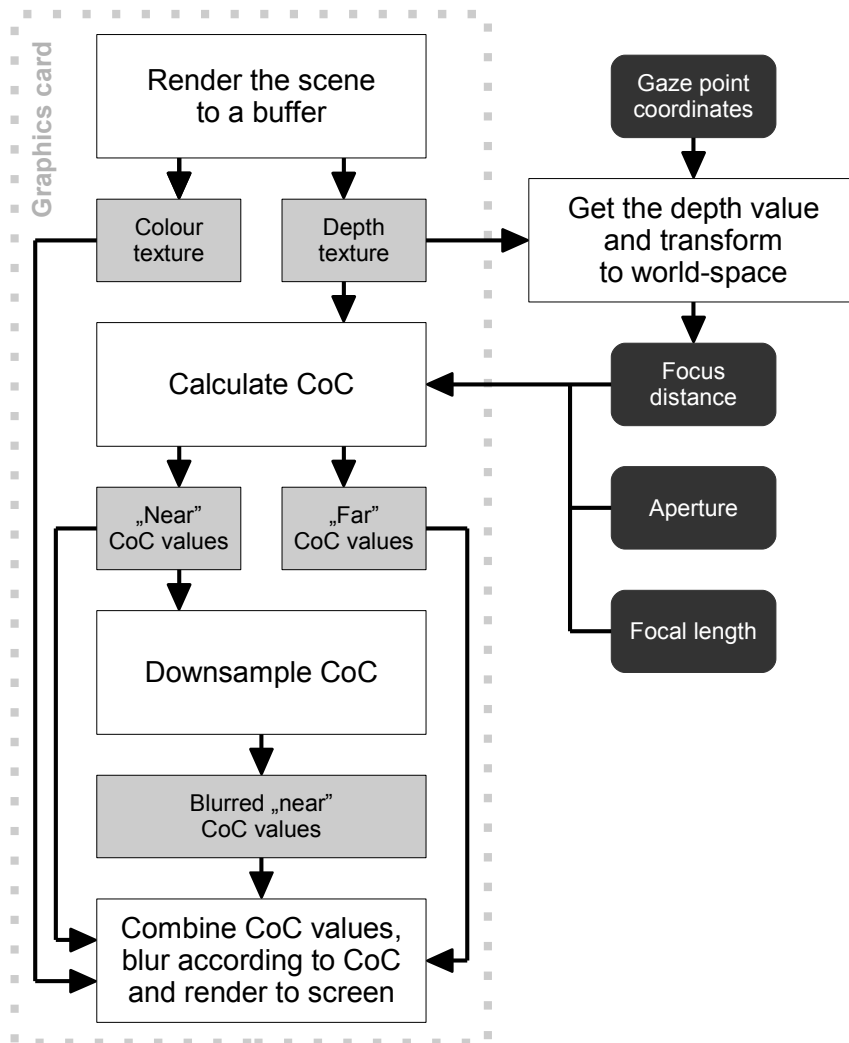
Celem testowej implementacji było utworzenie systemu, który dostarczałby obraz o akceptowalnej jakości przy zachowaniu płynności animacji. Przedmiotem prac nie było stworzenie zoptymalizowanej implementacji pozwalającej na syntezę obrazu z największą wydajnością.

### 3.1.3 Badanie skuteczności

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].

#### 3.1.3.1 Eksperyment

Zaimplementowane rozwiązanie było podstawą do przeprowadzenia eksperymentu percepcyjnego, mającego na celu weryfikację postawionej hipotezy. Przebadano 16 osób w wieku 21-24 lat, w tym jedną kobietę. Badane osoby oglądały animowane



Rysunek 3.12: Schemat przepływu danych w systemie interaktywnej symulacji zjawiska głębi ostrości. Ilustracja zaczerpnięta z autorskiej publikacji [MBT11]



Rysunek 3.13: Przykładowe klatki z animacji użytych podczas procedury eksperymentalnej.

sekwencje syntezowane w czasie rzeczywistym, przedstawiające wnętrze pomieszczenia (zob. Rys. 3.13). W niektórych sekwencjach wirtualna kamera pozostawała w spoczynku, w innych poruszała się. We wszystkich przypadkach symulowano zjawisko głębi ostrości w czterech wersjach: interaktywną, wykorzystującą na bieżąco dane okulograficzne z małym ( $a = 2$ ), średnim ( $a = 7$ ) i dużym ( $a = 19$ ) rozmyciem (zróżnicowanie stopni rozmycia uzyskano trzema wielkościami przysłony w modelu głębi ostrości) oraz nieinteraktywną ze średnim rozmyciem (*non-interactive*). Zadaniem uczestników badania było udzielenie odpowiedzi na pytanie: ”*Czy przedstawiona symulacja zjawiska głębi ostrości wyglądała realistycznie?*” za pomocą 5-stopniowej, dyskretnej skali Likerta (*Bardzo realistycznie, Realistycznie, Akceptowalnie, Nierealistycznie, Bardzo nierealistycznie*).

Sesja eksperymentalna każdego uczestnika rozpoczynała się od zapoznania się z pisemną instrukcją, po której następowała sesja treningowa pozwalająca zapoznać się z wirtualnym środowiskiem, z którego pochodziły wyświetlane później sekwencje testowe. Badani byli zaznajamiani ze sposobem działania okulografu. Właściwa część eksperymentalna zaczynała się od kalibracji systemu okulograficznego, po której następowała walidacja dokładności. Jeśli błąd celności obliczony w trakcie walidacji przekraczał 60 pikseli, a więc był większy niż  $1,5^\circ$  kąta widzenia na płaszczyźnie ekranu, eksperyment przerywano.

Badawcza część procedury eksperymentalnej oparta była o technikę *single stimulus*, gdzie obserwatorzy udzielali odpowiedzi po ekspozycji pojedynczego bodźca, a więc pojedynczej sekwencji z jedną z czterech wersji symulacji zjawiska głębi ostrości. Bodźce były wyświetlane w losowej kolejności, a każdy z nich był powtórzony

dwukrotnie, przy czym dbano o to by powtórzenia nie następowały bezpośrednio po sobie. By upewnić się, że analizowane dane będą tymi odpowiedziami, które zostały rzetelnie udzielone przez zaangażowanych już uczestników, na początku sesji wyświetlano dwa dodatkowe powtórzenia wybranych bodźców, bez zapisywania odpowiedzi udzielonych na następujące po nich pytania. Cała sesja trwała nie dłużej niż 8 minut, aby uniknąć zmęczenia badanych osób.

### 3.1.3.2 Wyniki

Podczas zbierania danych, skalę Likerta zakodowano kolejnymi liczbami całkowitymi, gdzie większe wartości oznaczają korzystniejszą ocenę realizmu (0 = *bardzo nierealistycznie*, 4 = *bardzo realistycznie*).

Wyniki przedstawiono na Rys. 3.14. Uczestnicy wskazywali nieinteraktywną symulację głębi ostrości jako mniej korzystną (wartość oczekiwana po standaryzacji  $z = -0,78$ , błąd standardowy  $sem = 0,11$ ). Jako najbardziej korzystną, badani wskazali wersję wykorzystującą na bieżąco dane okulograficzne i dokonującą średniego rozmycia ( $a = 7$ ) ( $z = 0,54$ ,  $sem = 0,07$ ). Inne rozmiary rozmycia spotkały się ze zbliżoną oceną do tego ostatniego.

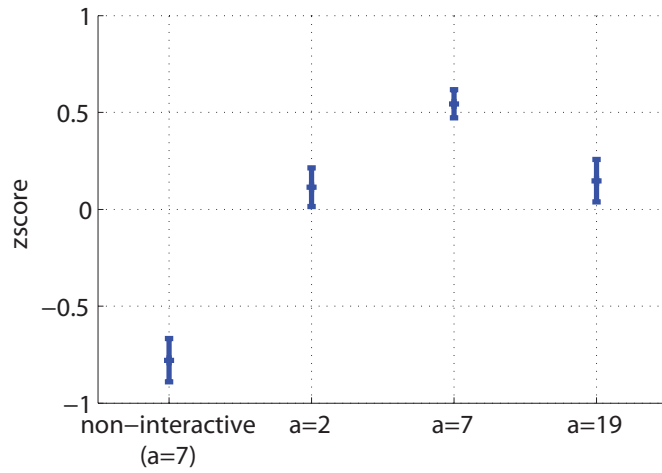
Analiza wariancji ANOVA nie wykazała istotnego wpływu poszczególnych scen ( $p = 0,48$ ) na wyniki. Nie znaleziono także statystycznie istotnej zależności od wyniku walidacji danych okulograficznych.

### 3.1.3.3 Wnioski

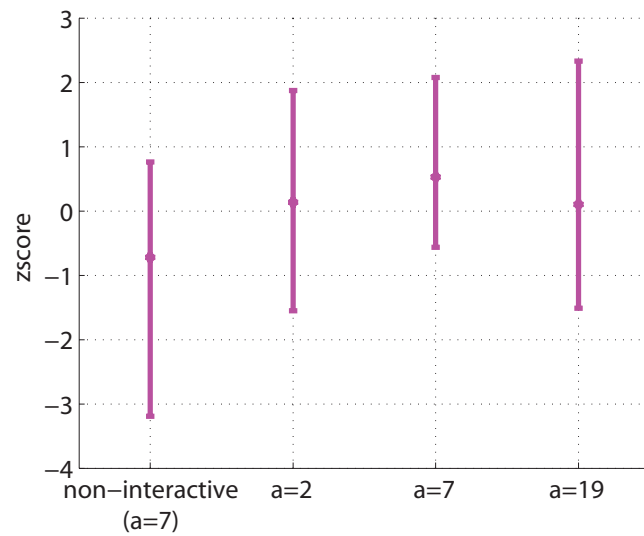
W eksperymencie operowano w dziedzinie iluzji pozorności *Psi* (ang. *Plausibility Illusion*, por. 1.1.1.2), zdając sobie sprawę z tego, że wywołanie iluzji miejsca *PI* (ang. *Place Illusion*) nie było możliwe przy zastosowanej technice syntezy obrazu i jego wyświetlania, nawet gdy głębia ostrości była symulowana w sposób interaktywny. Założenie, że badani mieli wrażenie znajdowania się w wirtualnej rzeczywistości, byłoby naiwne. Jednak uczestnicy eksperymentu mogli dokonać oceny oglądanego obrazu poprzez porównanie jego zachowania z tym, które znają jako rzeczywiste. Uzyskane odpowiedzi należy więc rozumieć jako ocenę naśladowania zachowania zjawiska głębi ostrości, a nie jego wiernego oddania.

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, naśladowując 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.

Hipoteza dotycząca wpływu interaktywnej symulacji głębi ostrości na postrzeganie realizmu i preferencję oglądanego obrazu może zatem zostać uznana za potwierdzoną w zakresie omawianego badania.



(a)



(b)

Rysunek 3.14: *Standaryzacja Z* (ang. *z-score*) udzielonych odpowiedzi. Rozpatrzono cztery przypadki: symulacja nieinteraktywnej głębi ostrości ze średnim zasięgiem rozmycia (*non-interactive*), interaktywna symulacja z małym rozmyciem ( $a = 2$ ), interaktywna symulacja ze średnim rozmyciem ( $a = 7$ ) oraz interaktywna symulacja z dużym rozmyciem ( $a = 19$ ). Wąsami na wykresie (a) przedstawiono błąd standardowy, na wykresie (b) bezwzględny zasięg otrzymanych wartości. Znacznikiem na obu wykresach oznaczano ustandaryzowaną wartość średnią. Ilustracje zaczerpnięte z autorskiej publikacji [MBT11]



## Dodatkowe obserwacje

Podczas udziału w eksperymencie, niektórzy z uczestników zwracali uwagę na zbyt szybkie zmiany rozmycia obrazu przy przenoszeniu wzroku na obiekt znajdujący się w innej odległości od kamery. Przy czym czas płynnej zmiany ostrości obrazu podczas interaktywnej symulacji ustalono na 370 milisekund, a więc był to czas zbliżony do typowego czasu akomodacji ludzkiego wzroku [Mil14]. Spostrzeżenia te mogły więc wynikać ze świadomości, że symulacja zjawiska nie jest procesem naturalnym, a więc oczekuje się jej wzmocnienia, przerysowania (jest to zjawisko, które wystąpiło też przy badaniu preferencji kontrastu, por. 3.2.2). Możliwym wytłumaczeniem jest też pojawiająca się w takiej sytuacji znacząca zmiana dużej części obrazu, szczególnie w przypadkach, gdy rozmycie było duże ( $a = 19$ ). Nawiązujący do czasu trwania naturalnej akomodacji oka, czas trwania tak rozległej zmiany obrazu, mógł wówczas powodować wrażenie szoku, co skutkowało takimi obserwacjami.

W późniejszej fazie eksperymentu, kilku uczestników zgłaszało problem, gdy prezentowane były bodźce z nieinteraktywną symulacją głębi ostrości. Kilkukrotna wcześniejsza możliwość korzystania z symulacji interaktywnej powodowała, że badani przyzwyczajali się do możliwości wpływania na odległość płaszczyzny ostrzenia poprzez przeniesienie wzroku na inny obiekt. Brak takiej możliwości stawał się frustrujący, co dodatkowo można zinterpretować na korzyść wpływu interaktywnej symulacji na komfort oglądania obrazu.

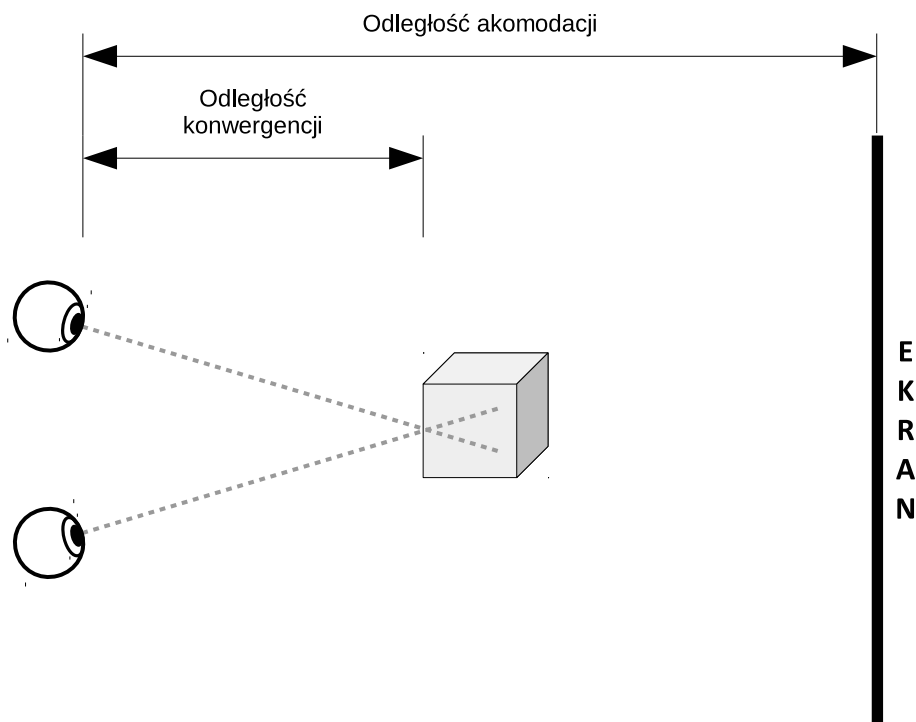
Istotną kwestią jest również istotność odpowiedniej metody wyznaczania odległości ostrzenia na podstawie wiedzy o punkcie skupienia wzroku. W eksperymencie zastosowano uproszczoną metodę odczytującą odległość od kamery pojedynczego piksela, po uprzedniej detekcji fiksacji metodą  $I-DT$  z poruszającym się w czasie oknem dyspersji (por. 2.2.4.1). Niektórzy uczestnicy badania zwrócili uwagę na to, że mieli utrudnioną możliwość spowodowania, że niewielkie (np. świecznik) lub szybko poruszające się przedmioty (np. nogi krzesła przy ruchu kamery przy podłodze) zostały wybrane jako te, na których ustawiona jest ostrość. Użycie lepszej metody wyboru obiektu uwagi, np. metody  $GDOT$  (zob. 2.3), mogłoby pomóc w tej kwestii.

### 3.1.4 Niezgodność konwergencji i 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. W tym podrozdziale opisano wykorzystanie interaktywnej symulacji głębi ostrości przy wyświetlaniu obrazu za pomocą techniki stereoskopii. Przeprowadzone badania zostały szerzej omówione na łamach współautorskiej publikacji [Duc14].

#### 3.1.4.1 Geneza problemu

Wyświetlanie obrazu stereoskopowego związane jest z zakłóceniem naturalnej zależności pomiędzy konwergencją i akomodacją. Poprzez wyświetlanie obrazów o nie-



Rysunek 3.15: 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.

zerowym 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. 3.15).

Problem ten, nazywany niezgodnością konwergencji i akomodacji (ang. *vergence-accommodation conflict*), jest uważany za podstawową przyczynę astenopii podczas oglądania obrazów stereoskopowych, a więc dyskomfortu wiążącego się ze zmęczeniem aparatu widzenia [LFH09, IKT09, HR02]. Zakłada się, że dyskomfort ten występuje wtedy, gdy dochodzi do konfliktowej reakcji aparatu widzenia, podyktowanej jedną ze wskazówek głębi, a stojącą w sprzeczności z informacją płynącą z innej wskazówki. Może się on objawiać bólami oczu i głowy obserwatora. Jest to przerwanie naturalnej korelacji obydwu mechanizmów [TSB11, RHFL10, SFW<sup>+</sup>11].

Z konfliktem związane jest pojęcie strefy komfortu (ang. *zone of comfort*) [TSB11]. Tym terminem nazywa się zakres różnic odległości ostrzenia od odległości konwergencji, który nie powoduje negatywnych odczuć. Im fizyczna odległość ekranu jest większa, tym większa jest tolerancja na różnicę pomiędzy rozsunieniem obrazów

obojga oczu, a niepoprawnym rozmyciem wynikającym z akomodacji.

W 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, miałyby więc spełnić w postrzeganiu głębi tę samą rolę, co naturalne rozmycie obrazu rzutowanego na siatkówkę.

#### 3.1.4.2 Określenie odległości

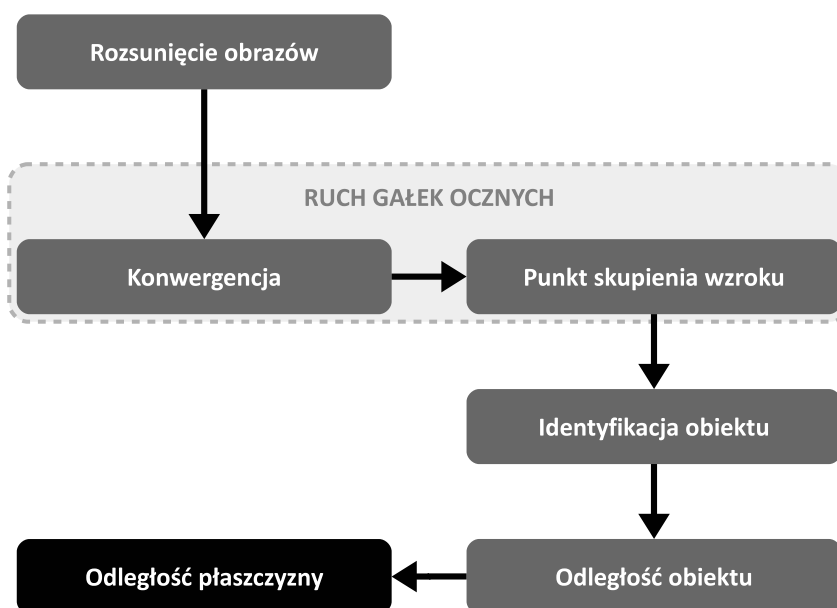
Dla realizacji tej koncepcji konieczne jest dążenie do jak najwierniejszego naśladowania takiego rozmycia na prezentowanym obrazie, które widziałby obserwator przy akomodacji do zadanej odległości. Wartością sterującą powinna więc być nie tyle odległość tego obiektu na ekranie, który odpowiadałby estymowanemu punktowi skupienia wzroku w przestrzeni wyświetlanej klatki, co fizyczna odległość, do której dostosowany jest w danej chwili aparat widzenia w rezultacie działania innych wskaźników głębi. Mając do dyspozycji okulary binokularny, a więc zdolny do niezależnej estymacji kierunku spojrzenia obojga oczu użytkownika, możliwe jest podjęcie próby określenia odległości płaszczyzny, do której za pomocą konwergencji dostosowany jest wzrok. Przyczyną samych ruchów konwergencyjnych może być na przykład celowo wprowadzone rozsuniecie obrazów prezentowanych obojgu oczom, typowe dla obrazowania stereoskopowego.

#### Odległość na podstawie konwergencji oczu

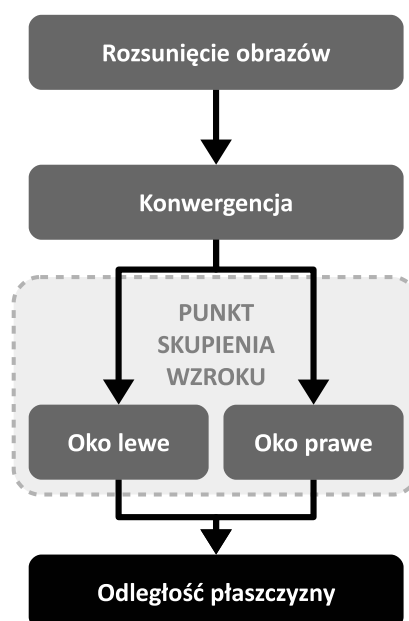
Estymacja odległości na podstawie pomiaru konwergencji jest mniej podatna na problem dualizmu w identyfikacji obiektu uwagi (por. Rozdz. 2.1.1), który dla tego zastosowania mógłby wprowadzić dodatkowe, bardzo niepożądane przekłamania. Gdy taka identyfikacja odbywa się na podstawie dopasowania uśrednionego, pojedynczego, wspólnego dla obojga oczu punktu skupienia wzroku do wyświetlanego obrazu obiektu, należy liczyć się z tym, że badana osoba może wcale nie skupiać się na postrzeganiu danego elementu. Takie zjawisko może być szczególnie częste, gdy w jego sąsiedztwie znajdują się obiekty bliższe lub dalsze obserwatora. Możliwe też, że odległość, na której skupia się w danej chwili wcale nie powinna odpowiadać odległości któregośkolwiek z wyświetlanych elementów. Przy błędnym rezultacie identyfikacji, otrzyma się więc niepoprawną odległość, która będzie rzutowała na dalszą błędność w działaniu użytego podejścia.

Natomiast biorąc pod uwagę odległość z konwergencji, przy założeniu doskonałego jej pomiaru, można pominąć ten problem poprzez dostosowanie symulowanej akomodacji do rzeczywistego stanu aparatu widzenia, bez wprowadzania dodatkowego etapu jakim byłoby odczytanie odległości z wirtualnej sceny (zob. Rys. 3.16). Jest to więc podejście preferowane.

Sygnal o odległości wynikającej z konwergencji sam mógłby również stać się dodatkowym, wspomagającym wejściem do systemu identyfikacji obiektu, na który skierowana jest w danej chwili uwaga obserwatora (por. Rozdz. 2.3). Przy trójwymiarowej scenie i obrazowaniu stereoskopowym, ułatwiłoby to rozróżnienie obiektów znajdujących się na różnych planach w prezentowanej wizualizacji.



(a)



(b)

Rysunek 3.16: Różnica w sposobie estymacji odległości płaszczyzny, na której skupiony jest wzrok użytkownika. W przypadku (a) na podstawie punktu skupienia wzroku identyfikowany jest znajdujący się na ekranie obiekt, a następnie znajduje jego odległość od ekranu. Jest ona przyjmowana za odległość, na której skupia się obserwator. W przypadku (b) odległość ta jest bezpośrednio wyliczana na podstawie osobnych punktów skupienia wzroku dla oka lewego i prawego.

Aby uzyskać odległość punktu, do którego dostosowana jest konwergencja oczu, zakładając ich horyzontalne położenie, wystarczy wziąć pod uwagę poziome współrzędne punktów skupienia wzroku obojga oczu. Ich rozsuniecie (ang. *disparity*), po uprzednim wyliczeniu mapowania na podstawie trójwymiarowej kalibracji, może posłużyć do obliczenia pożądanej odległości [Pfe12, EDP<sup>+</sup>12]. Tak uzyskany sygnał może być jednak bardzo silnie zaszumiony. Wskazane jest zastosowanie filtracji, np. używając filtra *Butterwortha* [DPHW11b].

### 3.1.4.3 Złagodzenie skutków niezgodności

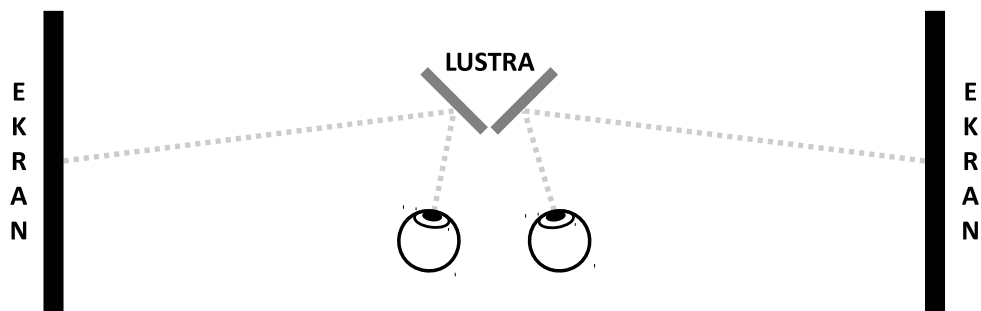
Za punkt wyjścia przyjęto wyniki badań zrealizowanych przez zespół Hoffmana [HGAB08], który opracował wielopłaszczyznowy wyświetlacz stereoskopowy (ang. *multi-focal display*). Celem prac zespołu było zbadanie redukcji dyskomfortu wynikającego z niezgodności konwergencji i akomodacji w sytuacji, gdy obraz stereoskopowy jest prezentowany na kilku płaszczyznach znajdujących się w różnych fizycznych odległościach od obserwatora. W ten sposób udało się uzyskać naturalnie występujące rozmycie na poszczególnych płaszczyznach, które według eksperymentalnie uzyskanych danych, doprowadziło do zmniejszenia wrażenia dyskomfortu u oglądających obraz osób. Dodatkowym skutkiem była poprawa szybkości i dokładności postrzegania głębi sceny. Stosując podobną metodologię badawczą, przeprowadzono eksperyment mający zweryfikować skuteczność interaktywnej symulacji rozmycia wynikającego z głębi ostrości, opartej o jednopłaszczyznowy wyświetlacz stereoskopowy i zbierane na bieżąco dane okulograficzne.

## Środowisko badań

Badanie przeprowadzono wykorzystując wyświetlacz stereoskopowy (ang. *haploscope*) składający się z dwóch oddzielnych ekranów, a więc implementujący ideę nazywaną stereoskopem Wheatstone'a (por. Rozdz. 1.1.2.1). Każdy z nich prezentuje obraz dla jednego z oczu użytkownika. Tor widzenia kierowany jest na odpowiedni dla danego oka wyświetlacz za pomocą luster (zob. Rys. 3.17). Zastosowanie takiego rozwiązania pozwoliło uniknąć stosowania technik wyświetlania stereoskopowego polegających na wykorzystaniu dodatkowych elementów w torze optycznym, takich jak okulary migawkowe lub filtry barwne. Elementy te mogłyby utrudnić wykorzystanie systemów okulograficznych, a także samym swoim charakterem działania doprowadzić do zmęczenia wzroku.

Użyte wyświetlacze to ekrany wysokiej rozdzielczości *IBM T221*, co pozwala na wyświetlenie pełnej klatki w rozdzielczości  $3840 \times 2400$  pikseli dla każdego oka. Klatki generowane były przez system komputerowy wyposażony w dwie karty graficzne *NVIDIA Quadro FX 5800*. Długość toru optycznego od oka do ekranu wynosiła  $86\text{cm}$ , użyte lustra ustawione były pod kątem  $45$  stopni do płaszczyzny każdego z ekranów.

Za system okulograficzny posłużyły dwie kamery *LC Technologies* umieszczone poniżej wyświetlaczy i skierowane na odbicie oczu obserwatora w lustrach. Za przetwarzanie danych w celu uzyskania położenia punktów skupienia wzroku odpowiadał system *Eyegaze System*.



Rysunek 3.17: 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 lusterek.

### Procedura eksperymentalna

Przebadano 19 osób (10 kobiet i 9 mężczyzn), których średni wiek wynosił 24 lata (odchylenie standardowe 5,75). Uczestnicy brali udział w eksperymencie dwukrotnie, raz z włączoną symulacją głębi ostrości, a drugi raz bez niej. Podejścia odbywały się z jednodniowym odstępem. Próbę podzielono na dwie grupy o zbliżonej liczebności, każda z nich odbywała podejścia zaczynając od innego stanu.

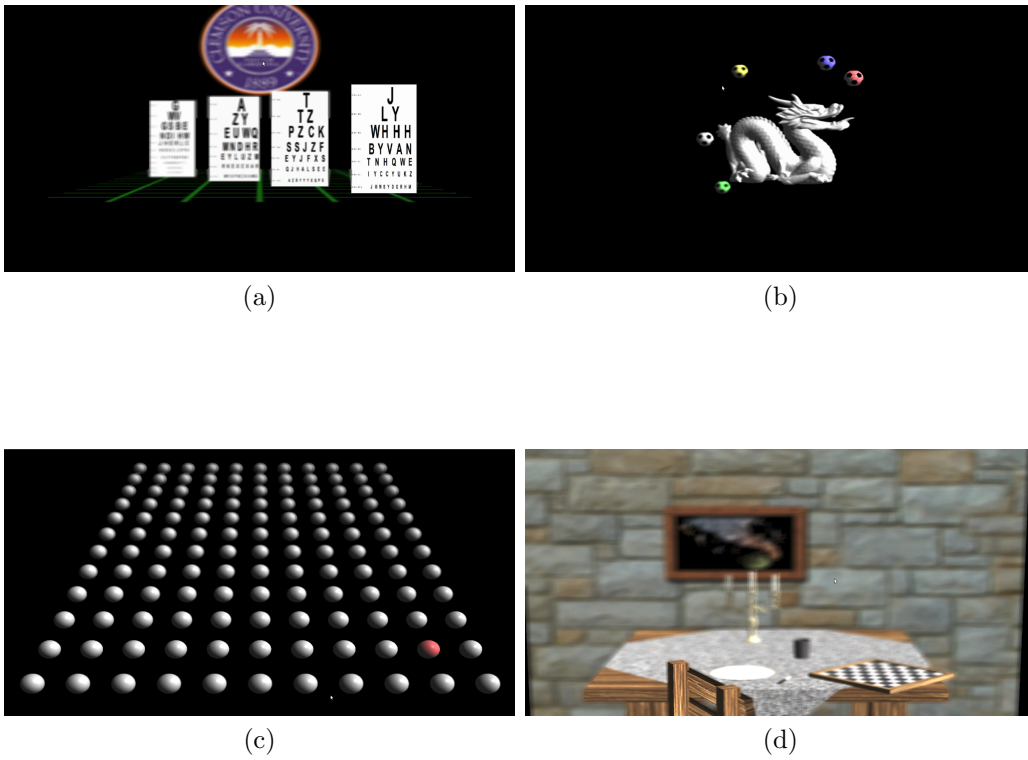
Każdorazowo przed rozpoczęciem eksperymentu, dokonywano trójwymiarowej kalibracji niezbędnej do późniejszej estymacji odległości, na której skupiany jest wzrok na podstawie binokularnych danych.

Uczestnicy mieli za zadanie przyglądać się czterem trójwymiarowym scenom, na każdej z których umieszczono obiekty znajdujące się w zróżnicowanych odległościach od obserwatora (zob. Rys. 3.18). Podczas oglądania każdej z nich, badanych proszono o spoglądanie w z góry narzucone miejsca. Dla sceny (a) był to jeden z wierszy każdej z tablic. Dla sceny (b) były to kolejno piłki o różnych kolorach. Dla sceny (c) oznaczano kolejno kilka spośród wyświetlonych piłek. Dla sceny (d) uczestnikowi werbalnie nakazywano patrzeć na poszczególne przedmioty w z góry ustalonej kolejności.

Po obejrzeniu scen, uczestnik miał za zadanie odpowiedzieć na zestaw pięciu pytań dotyczących jego odczuć. Odpowiedzi udzielane były z użyciem 100-punktowej skali będącej rozszerzeniem skali Likerta. Takie rozwiązanie miało za zadanie zwiększyć czułość na niewielkie różnice w ocenie subiektywnych wrażeń przez badane osoby.

Treść pytań starano się dobrać tak, by nie odwoływać się do pamięci uczestników na temat poprzedniego podejścia, a więc nie zmuszając ich do własnoręcznego dokonywania porównań. Była to, podobnie jak użycie 100-stopniowej zamiast 5-stopniowej skali, różnica metodologii w stosunku do użytej w badaniu zespołu Hoffmana [HGAB08]. Chcąc zbadać subtelne różnice przy wrażeniach odczuwanych przez badane osoby, zdecydowano się na pozostawienie im większej dowolności, jednocześnie unikając zmuszania ich do dokonania oceny porównawczej.

Zadano następujące pytania:



Rysunek 3.18: Sceny użyte w eksperymencie. (a) tablice Snellena umieszczone w różnej odległości od kamery, (b) smok z wirującymi wokół niego piłkami, (c) siatka piłek i (d) wnętrze pomieszczenia nawiązujące do sceny opisanej w Rozdz. 3.1.3.

- $Q_1$ : Czy odczuwasz podrażnienie oczu? (*zdecydowanie nie - zdecydowanie tak*)
- $Q_2$ : Czy odczuwasz ból głowy? (*zdecydowanie nie - zdecydowanie tak*)
- $Q_3$ : Czy odczuwasz zmęczenie? (*zdecydowanie nie - zdecydowanie tak*)
- $Q_4$ : Czy podobała ci się sesja? (*zdecydowanie nie - zdecydowanie tak*)
- $Q_5$ : Czy chciałbyś zobaczyć więcej obrazów tego typu? (*zdecydowanie nie - zdecydowanie tak*)

## Wyniki badania i ich analiza

Dla zmiennej zależnej w postaci odczuwalnego zmęczenia (pytanie  $Q_3$ ), udało się wykazać statystycznie istotną zależność wskazywanych wartości od faktu wystąpienia symulacji zjawiska głębi ostrości (zob. Rys. 3.19a). Uczestnicy oceniali swoje zmęczenie jako mniejsze, gdy dokonywano rozmycia elementów obrazu znajdujących się bliżej lub dalej od odległości wynikającej z konwergencji oczu (średnia ocena  $M = 5,95$ , błąd standardowy  $sem = 1,71$ ), niż gdy go nie dokonywano ( $M = 18,53$ ,  $sem = 4,13$ ). Przy analizie ANOVA uzyskano statystykę  $F = 6,34$  przy poziomie istotności  $p < 0,05$ ,  $\eta^2 = 0,185$ .

Rozpatrując występowanie bólu głowy (pytanie  $Q_2$ , zob. Rys. 3.19b), uczestnicy istotnie częściej wskazywali na jego występowanie w przypadku pozbawionym symulacji głębi ostrości ( $M = 29,63$ ,  $sem = 4,36$ ), niż gdy symulowano rozmycie ( $M = 2,84$ ,  $sem = 1,71$ ). Obliczono wartość statystyki  $F = 7,61$  przy poziomie istotności  $p < 0,02$  i  $\eta^2 = 0,168$ .

Dla pozostałych z pytań nie udało się wykazać statystycznie istotnej zależności.

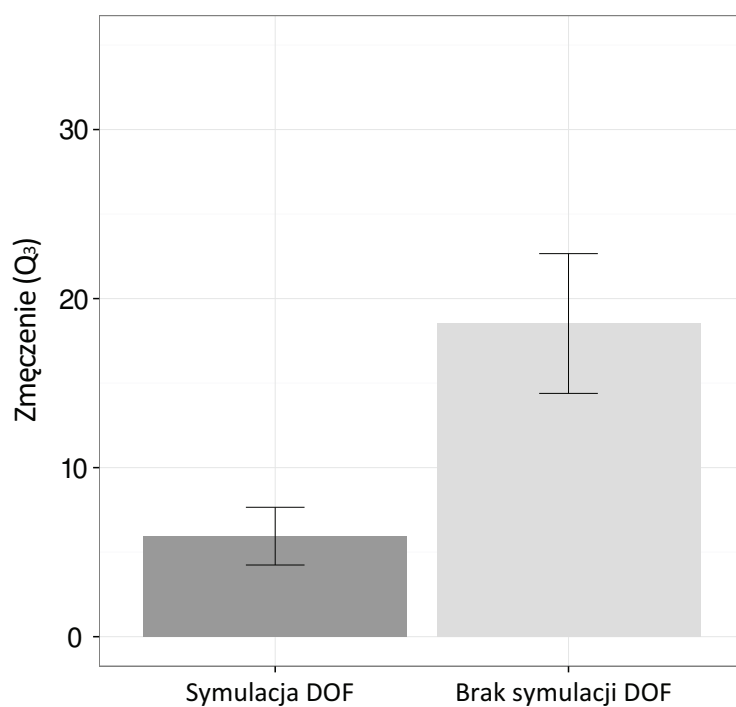
Na podstawie uzyskanych wyników można więc stwierdzić, że udało się potwierdzić hipotezę o łagodzeniu skutków konfliktu akomodacji i konwergencji, gdy prezentacji obrazu stereoskopowego będzie towarzyszyła symulacja rozmycia zależnego od odległości obiektu od płaszczyzny, na której skupiony jest wzrok obserwatora.

## 3.2 Głębia w obrazach o poszerzonym zakresie dynamiki luminancji

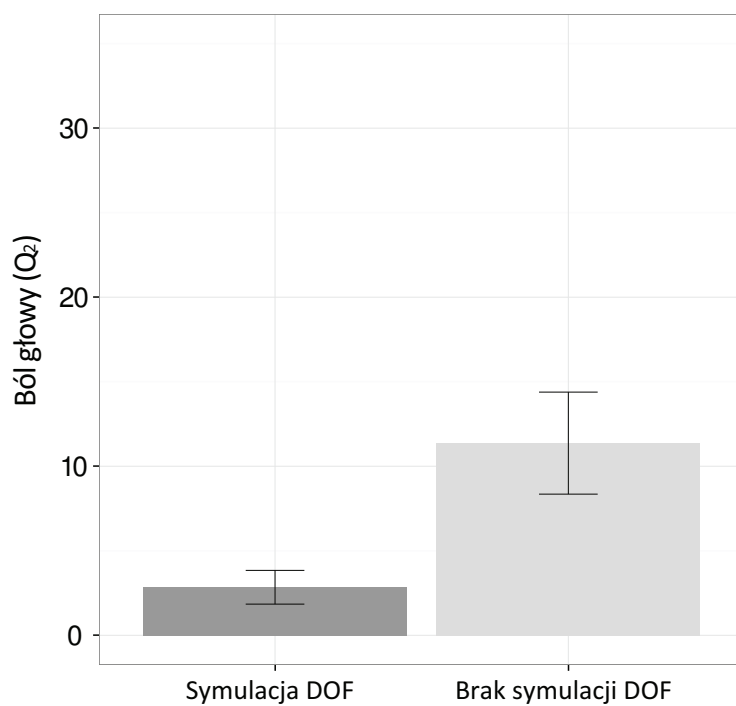
Pozostając w temacie wpływu syntetycznych czynników związanych z prezentacją obrazu na wrażenie głębi u obserwatora, postanowiono przyjrzeć się również kolejnej po głębi ostrości, monokularnej wskazówce: kontrastowi. W kontekście zyskujących popularność wyświetlaczy HDR (ang. *High Dynamic Range*, poszerzony zakres dynamiki luminancji), wykorzystanie oferowanych przez nie możliwości dla lepszego przekazywania informacji o zróżnicowaniu głębokości na scenie lub uzupełniania informacji zawartej w pozostałych wskazówkach, zdaje się być obszarem wartym zbadania.

Problem podzielono na dwie części: *wpływ kontrastu na widzenie głębi* oraz *wpływ kontrastu na wrażenie realizmu*. W dalszej części podrozdziału opisano badania, które były przedmiotem publikacji powstałej w ramach prac nad niniejszą rozprawą [Pet14]. W ich wyniku zamodelowano relację pomiędzy kontrastem a estymacją





(a)



(b)

Rysunek 3.19: Zobrazowanie uzyskanych rezultatów dla pytań  $Q_3$  (a) oraz  $Q_2$  (b). Słupki oznaczają średnią wartość dla wszystkich uczestników, wąsy natomiast pokazują błąd standardowy (ang. *Standard Error of Mean*).

głębi dla wyświetlaczy stereoskopowych o poszerzonym zakresie dynamiki luminancji oraz wskazano charakterystykę wpływu kontrastu i stereoskopii na postrzeganie obrazu jako bardziej zbliżonego do rzeczywistości.

### 3.2.1 Wpływ kontrastu na postrzeganie głębi

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*, myślą go z wyświetlaczem przedstawiającym obraz trójwymiarowy. Równie często pojawia się porównanie tak wyświetlanego obrazu ze spoglądaniem na rzeczywistą scenę przez okno [RHM11]. Weryfikacja hipotezy pozwoli na określenie, czy faktycznie da się wykazać mierzalną zależność postrzeganej odległości od kontrastu i luminancji, czy też wrażenia użytkowników zależą od interpretacji obrazu o poszerzonej dynamice jako bardziej realistycznego, a iluzoryczna przestrzenność nie przekłada się na rozróżnialną, możliwą do pomierzenia odległość.

W przypadku pomyślnej weryfikacji hipotezy, możliwe byłoby rozważenie zastosowania pozyskanej wiedzy do wzmocnienia informacji o głębi płynącej z prezentacji obrazu stereoskopowego. W ten sposób, w celu wytworzenia u oglądającego obraz takiego samego poczucia różnic odległości, można próbować ograniczyć rozsuniecie obrazów dla oka lewego i prawego, kompensując to ograniczenie odpowiednim wykorzystaniem zależności związanej z kontrastem. Dzięki temu możliwe byłoby złagodzenie konfliktu konwergencji-akomodacji poprzez pozostanie w strefie komfortu (por. Rozdz. 3.1.4).

Zaś w przypadku gdyby hipoteza okazała się fałszywa, a złudzenie przestrzenności dwuwymiarowych obrazów o poszerzonej dynamice nie niesłoby w sobie żadnej mierzalnej i rozróżnialnej informacji o głębi, podejście takie można by wykorzystać jako swojego rodzaju zastępczy mechanizm prezentacji trójwymiarowych scen na płaskim medium. Mechanizm wystarczający, gdy nie jest konieczne wprowadzenie faktycznych wrażeń dotyczących mierzalnych różnic w odległościach, a jedynie ogólnego odczucia przestrzenności pokazywanego obrazu.

#### 3.2.1.1 Klasyczne wyświetlacze

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 (por. Rozdz. 1.1.2).

Dla dalszych rozważań istotne jest przyjęcie użytych w tej pracy określeń, takich jak kontrast obszarowy (ang. *area contrast*) i kontrast tekstury (ang. *texture contrast*). Ten pierwszy rozumiany jest jako kontrast pomiędzy obiektem a jego tłem lub innym obiektem. Natomiast kontrastem tekstury nazywamy kontrast wewnątrz

obrazu danego obiektu, wynikający z jego tekstury lub wpływu oświetlenia. Obserwacją zespołu Ichihary było współdziałanie obu z nich dla tworzenia efektywnego wrażenia głębi.

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.

W wymienionych pracach badano jednak kontrast jako wyizolowaną wskazówkę głębi. Przedmiotem opisanych w tym rozdziale badań będzie analiza wpływu kontrastu na postrzeganie głębi w sytuacji, gdy obecne są także, uznawane za silniejsze, wskazówki binokularne w postaci obrazu stereoskopowego.

### 3.2.1.2 Wyświetlacze stereoskopowe

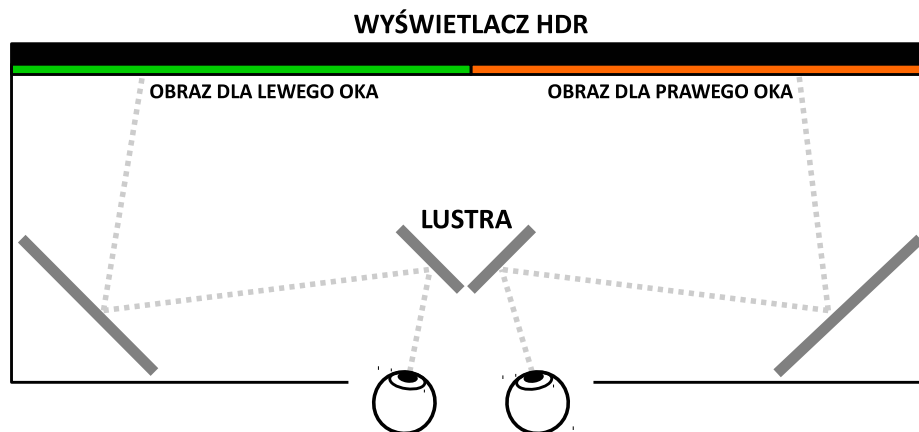
W eksperymentach wykorzystano 47-calowy wyświetlacz *SIM2 HDR47E*, charakteryzujący się zdolnością wyświetlania obrazu o rozdzielczości  $1920 \times 1080$  pikseli. Podświetlenie ekranu ciekłokrystalicznego, złożone z indywidualnie sterowanych bloków *LED*, pozwala na uzyskanie luminancji na poziomie  $4000 \text{ cd/m}^2$  przy statycznym kontraście  $20000 : 1$ .

Wyświetlacz ten stał się podstawą do budowy stereoskopu opartego o ideę Wheatstone'a. Ekran został logicznie podzielony na dwie części. Lewa połowa jest przeznaczona do wyświetlania klatek dla lewego oka, natomiast prawa połowa ekranu wyświetla klatki dla oka prawego. Za przekierowanie toru optycznego obserwatora na odpowiednie części ekranu odpowiada system czterech lusterek. Długość toru optycznego, a więc odległość pomiędzy oczami obserwatora a fizyczną powierzchnią ekranu, wynosi po uwzględnieniu odbić  $87.3 \text{ cm}$ . Całość zamknięta jest w obudowie posiadającej jeden otwór, który pozwala użytkownikowi obejrzeć prezentowany obraz (zob. Rys. 3.20 i Rys. 3.21).

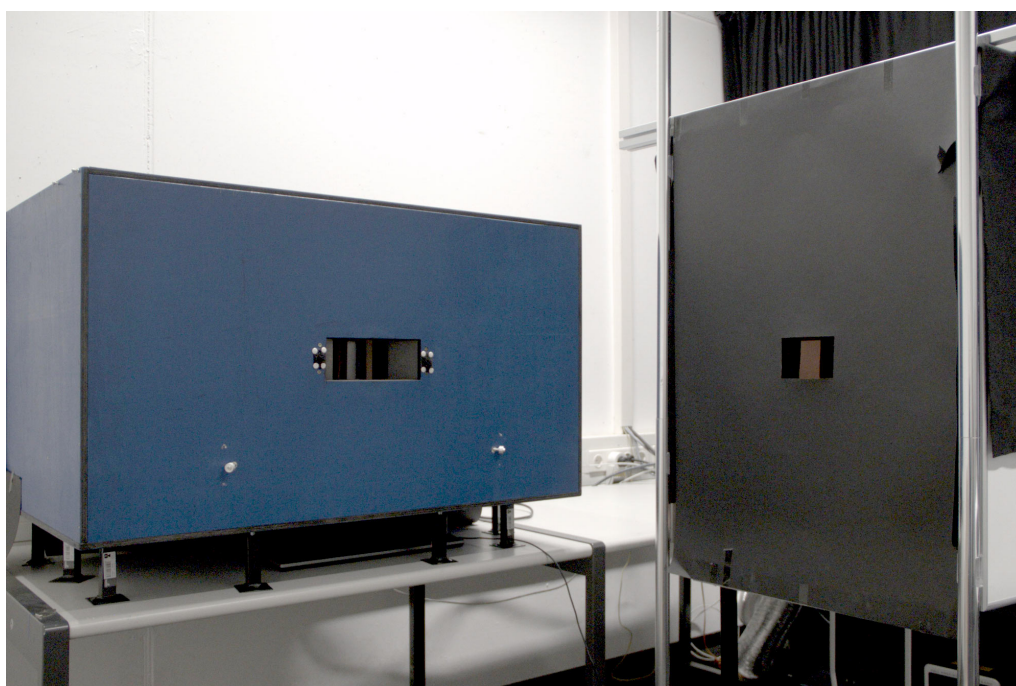
Zadbano o kalibrację jednorodności luminancji wyświetlanego obrazu z punktu widzenia obserwatora, by wyeliminować niedokładność samego wyświetlacza, jak i toru optycznego, w skład którego wchodziły lustra. Dokonano pomiaru luminancji obrazu za pomocą miernika *Konica Minolta LS-100* w  $16 \times 9$  punktach dla pięciu wartości luminancji ( $0.1, 1, 10, 100$  i  $1000 \text{ cd/m}^2$ ), które zadano wyświetlać monitorowi. Pomiar był dokonywany poprzez lustra stereoskopu. W ten sposób uzyskano mapę korekcji dla całej powierzchni wyświetlacza, którą stosowano do kompensacji niedokładności podczas prezentacji klatek w trakcie trwania eksperymentu. Korekcja następowała osobno dla każdego piksela wyświetlanego obrazu.

### 3.2.1.3 Eksperyment

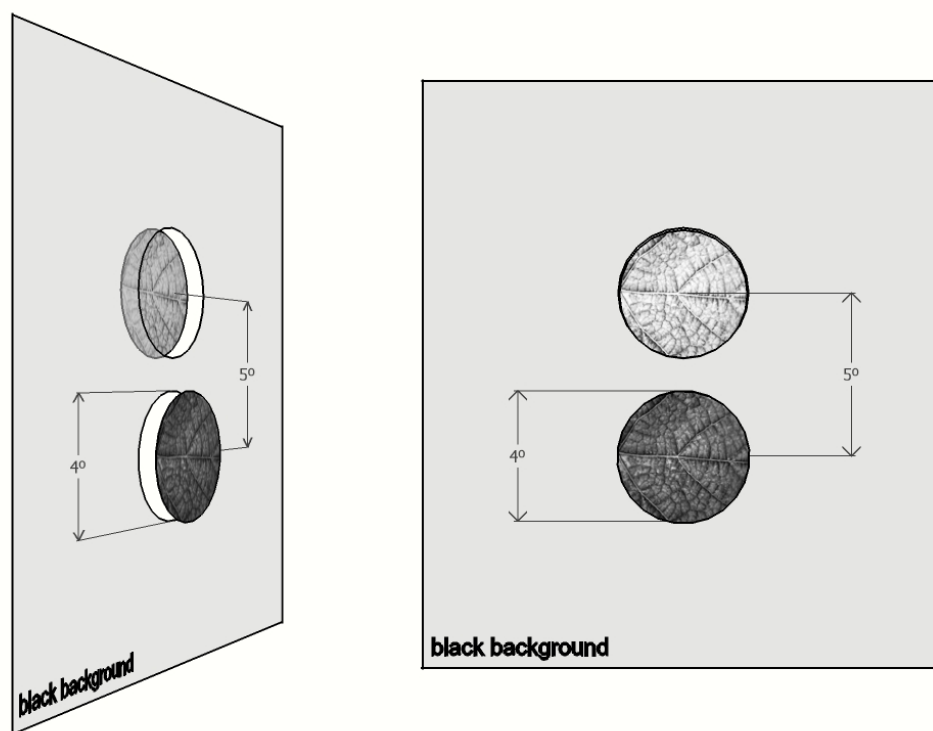
Za punkt wyjścia przyjęto badanie wykonane przez zespół Rempela [RHM11]. Odtworzono jego przebieg korzystając z opisanego wyżej środowiska testowego, uzyskując potwierdzenie wyników dotyczących działania kontrastu luminancji jako wskazówki głębi w przypadku braku innych przesłanek. Następnie wzbogacono ekspery-



Rysunek 3.20: 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 3.21: 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 Rozdz. 3.2.2.



Rysunek 3.22: Schematyczne przedstawienie bodźców, które zostały wykorzystane podczas badania wpływu kontrastu na postrzeganie głębi na wyświetlaczu stereoskopowym. Bodźce składały się z dwóch dysków o zmiennej średniej luminancji oraz zmiennej odległości wirtualnej płaszczyzny, na której były one umieszczone. Po lewej stronie przedstawiono wizualizację różnych odległości płaszczyzn, na których znajdują się poszczególne dyski. Oznaczone wymiary dane są w stopniach kąta widzenia obserwatora.

ment o wskazówki binokularne wynikające z rozsunienia obrazów oka lewego i prawego.

### Bodźce

Podczas badania uczestnicy oglądali prezentowane na stereoskopowym wyświetlaczu *HDR* bodźce składające się z par dysków, ułożonych jeden pod drugim. Każdy z dysków był wypełniony możliwie neutralną teksturą powierzchni liścia, o rozkładzie jasności pikseli zbliżonym do rozkładu normalnego i kontraście tekstury równym 0.5. Użyta tekstura była identyczna jak w badaniu przeprowadzonym przez zespół Rempela [RHM11]. Tło bodźca było jednolite, zawsze o luminancji równej  $1 \text{ cd/m}^2$  (zob. Rys. 3.22).

Kolejno wyświetlane bodźce różniły się od siebie średnią luminancją dysków oraz ich rozsunieniem pomiędzy klatką dla oka lewego i prawego, co odpowiadało różnym odległościom płaszczyzn od obserwatora, na których miałyby się one znajdować.

Przyjęto dwa poziomy średniej luminancji dla wyświetlanych dysków: *ciemny* ( $50 \text{ cd/m}^2$ ) i *jasny* ( $1000 \text{ cd/m}^2$ ). Prezentowano bodźce składające się z dwóch

dysków, ukazujące kombinacje *ciemny-ciemny*, *jasny-jasny* oraz mieszane *ciemny-jasny*. Kolejność ich pokazywania, rozumiana zarówno jako kolejność bodźców jak i kolejność samych dysków w obrębie bodźca, była losowa.

Badając działanie kontrastu jako modyfikatora wrażenia odległości przy jednoczesnym występowaniu silnej wskazówki jaką jest rozsuniecie stereoskopowe, spodziewano się zaobserwować różnicę jedynie dla warunków bliskich progowi rozróżnialności odległości (ang. *near-threshold*). Użyto czterech odległości wirtualnych płaszczyzn na których umieszczano dyski: 86.7, 86.9, 87.1 i 88.2 cm, z czego trzy pierwsze znajdowały się przed rzeczywistą płaszczyzną ekranu, a ostatnia za nią. Żaden z dysków nie był przedstawiany w tej samej odległości od obserwatora, co fizyczny ekran.

Podczas eksperymentu wyświetlano wszystkie kombinacje par wymienionych odległości. W efekcie otrzymano bodźce, w których różnica rozsunieć obrazów oka lewego i prawego pomiędzy prezentowanymi dyskami wynosiła  $0'$ ,  $\pm 0.43'$ ,  $\pm 0.85'$ ,  $\pm 3.41'$ ,  $\pm 3.83'$  lub  $\pm 4.26'$ , gdzie  $'$  oznacza minuty kąta widzenia obserwatora.

Prezentacja bodźców była rozdzielona pokazaniem przez pół sekundy pustego ekranu o standardowej luminancji tła bodźca.

## Uczestnicy

W badaniu wzięło udział 11 osób, w tym 2 kobiety i 9 mężczyzn. Średnia wieku badanych to 23.4 lat. Wśród uczestników 5 osób nosiło podczas badania okulary korekcyjne. Każda z biorących udział w eksperymencie osób deklarowała poprawne widzenie stereoskopowe i nie była świadoma celu badania.

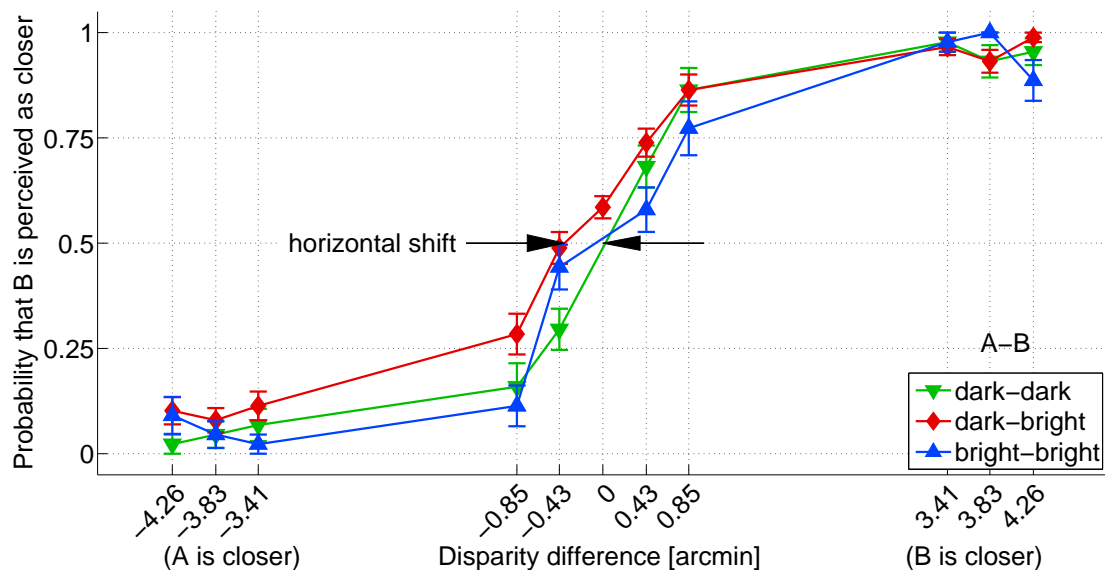
## Procedura badawcza

Badani mieli za zadanie wybierać za pomocą klawiatury, który z prezentowanych w danym bodźcu dysków znajduje się na bliższej płaszczyźnie. Udzielenie odpowiedzi w postaci jednej z dwóch możliwości było obowiązkowe, eksperyment miał więc budowę *forced choice*. Pięć pierwszych pokazanych bodźców miało charakter sesji treningowej, obejmującej reprezentatywny zakres kontrastów i rozsunieć użytych później we właściwej części. Łączna liczba pokazywanych bodźców, wyłączając treningowe, wynosiła 220 dla każdego z uczestników (8 powtórzeń każdej z 28 kombinacji). Czas trwania badania nie był ograniczony, każdy uczestnik miał dowolność w ustaleniu tempa przechodzenia do kolejnych bodźców.

### 3.2.1.4 Wyniki

Rezultat badania w postaci uśrednionych funkcji psychometrycznych dla trzech przypadków średnich luminancji dysków (*ciemny-ciemny*, *jasny-jasny* i *ciemny-jasny*), przedstawiono na Rysunku 3.23. Dla sytuacji, w których poziomy luminancji dysków zgadzają się ze sobą, postrzegana odległość jest zgodna z faktycznie wyświetlanymi rozsunieniami. Dla sytuacji w której jeden z dysków jest *jasny*, a drugi *ciemny*, daje się zauważyć tendencja do wskazywania jaśniejszego z nich jako bliższego (przesunięcie zaznaczone strzałkami na Rysunku 3.23).

Wykazana zależność odpowiada różnicy odległości na poziomie 1.4 mm przy różnicy luminancji między poziomami *jasnym* i *ciemnym*, a więc dla stosunku lu-



Rysunek 3.23: Wykres funkcji psychometrycznych będących rezultatem eksperymentu badającego zależność postrzeganej różnicy odległości od kontrastu w warunkach stereoskopowych. Pozioma oś przedstawia różnice pomiędzy rozsunięciami par dysków (dane w minutach kąta widzenia). Lewa strona diagramu prezentuje sytuację, w której dysk oznaczony jako *A* jest faktycznie bliższy obserwatora, zaś prawa strona sytuację gdy dysk *B* jest bliższy. Na osi pionowej oznaczono prawdopodobieństwo wybrania przez uczestnika dysku *B* jako bliższego. Kolorami oznaczono trzy przypadki średnich luminancji wyświetlanych dysków.

minancji 200 : 1. Jest to więc zależność o niewielkim zasięgu i zanika ona, gdy wskazówka binokularna jest odpowiednio silna, a więc znacząco powyżej progu jej wykrywalności.

### 3.2.1.5 Wnioski

Pomimo potwierdzenia postawionej wcześniej hipotezy, 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 rozsunięcia klatek dla obojga oczu wiązałoby się z wyjściem poza strefę komfortu widzenia stereoskopowego. Nie należy również zapominać, że w przypadku braku występowania innych wskazówek, luminancja i kontrast stanowią bardzo silną przesłankę o szeregowaniu obiektów pod względem odległości od obserwatora [EML13].

## 3.2.2 Kontrast i głębia a poczucie realizmu

Chcąc poznać przyczynę raportowanego przez zespół Rempela silnego poczucia realności obrazów oglądanych na wyświetlaczu *HDR* [RHM11], przygotowano eks-

peryment bazujący na obrazach rzeczywistej sceny. 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.

### 3.2.2.1 Eksperyment

W opisanym badaniu realizm zostanie rozpatrzony jako wielkość psychofizyczna, będąca funkcją dwóch atrybutów: kontrastu prezentowanego obrazu oraz dystansu pomiędzy osiami widzenia obojga oczu. Kontrast będzie dany wartością wykładnika  $\gamma$ , który posłuży do modyfikacji wartości luminancji zarejestrowanych fotografii (wytłumaczenie jego zastosowania znajduje się w dalszej części tego podrozdziału). Zaś dystans pomiędzy osiami widzenia będzie określany jako *IAD*, co odpowiada angielskiemu *Inter-Axial Distance* i, zależnie od rozpatrywanego przypadku, nawiązuje do rozstawu oczu obserwatora lub osi optycznych obiektywów, z użyciem których fotografuje się scenę, a później rozsunięcia obrazów przeznaczonych dla oka lewego i prawego. Im większa wartość *IAD*, tym sugestia dużych różnic odległości pomiędzy obiektami powinna być większa. Można spodziewać się, że wartość *IAD* równa rzeczywistemu rozstawowi oczu obserwatora powinna skutkować najbardziej naturalnymi wrażeniami.

Opisane dalej badanie zostało przeprowadzone korzystając ze sprzętu opisanego wcześniej w Rozdziale 3.2.1.2. Dodatkowo na potrzeby tego eksperymentu przeprowadzono kalibrację koloru wyświetlanego obrazu, poprzez iteracyjny proces pokazywania obrazu na wyświetlaczu, fotografowania rezultatu z punktu widzenia obserwatora i kolejnego wyświetlania obrazu - tym razem zastąpionego wykonaną w poprzednim kroku fotografią. Przy kolejnych iteracjach dokonywano korekty kolorów poprzez obliczenie i zastosowanie podczas wyświetlania macierzy transformacji minimalizującej odległości kolorów pikseli w przestrzeni *CIE 1976* ( $L^*$ ,  $u^*$ ,  $v^*$ ). W ten sposób różnica kolorów dążyła do zera. Celem tej kalibracji było ujednoczenie kolorów pomiędzy rzeczywistą sceną, a wyświetlaczem.

### Bodźce

By móc zweryfikować hipotezę, konieczne było posłużenie się fotorealistycznymi obrazami. W tym celu przygotowano zamkniętą w obudowie scenę składającą się z różnych elementów, których powierzchnie były zarówno matowe, błyszczące jak i przezroczyste. Tak, by objąć możliwie wiele spotykanych w rzeczywistości sytuacji, mogących dostarczać obserwatorowi dodatkowych przesłanek stanowiących o badanym odczuciu realizmu.

Obiekty znajdowały się na różnych odległościach od przyszłego obserwatora, w zakresie od 65 do 150 cm. Zakres ten odpowiada strefie komfortu związanej z konfliktem konwergencji-akomodacji, gdy docelowe obrazy mają być wyświetlane na ekranie oddalonym o 87.3 cm od oczu. Na scenie umieszczone było pojedyncze źródło światła (zob. Rys. 3.24 oraz Rys. 3.21).





Rysunek 3.24: 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.

Wykonano serię fotografii stereoskopowych, umieszczając aparat fotograficzny na statywie pozwalającym na kontrolę odległości pomiędzy miejscami z których zostaną wykonane zdjęcia dla punktu widzenia oka lewego i prawego, co wprost przekładało się na wartość  $IAD$ . Należy też nadmienić, że zdjęcia były wykonywane z miejsca odpowiadającego położeniu oczu późniejszego obserwatora sceny. Użyto  $IAD$  równych 0, 3.25, 6.5 oraz 9.75 cm. Wartość zerowa odpowiada brakowi stereoskopii (ta sama klatka dla oka lewego i prawego), natomiast wartość 6.5 cm odpowiada przeciętnemu rozstawowi oczu u dorosłego człowieka.

Dla każdej z wymienionych wartości  $IAD$  wykonano serię zdjęć z różnymi ustawieniami ekspozycji zarówno dla oka lewego, jak i prawego (wyjątkiem jest wartość 0 cm, gdzie nie wykonywano oddzielnych zdjęć dla obojga oczu). Pokryto czasy naświetlania od 1/4000 do 30 s z odstępami wielkości 1- $f$ . Przysłona była ustawiona na wartość  $f/8$ , by uniknąć krótkiej głębi ostrości mogącej rozmyć niektóre z obiektów sceny. Ogniskowa obiektywu została dobrana tak, by pokrycie rzeczywistej sceny odpowiadało obszarowi na którym będzie wyświetlany obraz. Czułość  $ISO$  ustawiono na wartość 200 by uniknąć zaszumienia.

Serie fotografii zostały połączone w obrazy  $HDR$  z użyciem oprogramowania *pfstools* [MKMS07] oraz krzywej reprodukcji luminancji charakterystycznej dla użytego aparatu fotograficznego. W efekcie otrzymano klatki zawierające informację o luminancji odpowiadającą rzeczywistym wartościom.

Zróznicowany kontrast na potrzeby badania uzyskano poprzez skalowanie luminancji poszczególnych pikseli z użyciem globalnego operatora tonów  $gamma$ :

$$Y_{out} = \left( \frac{Y_{in}}{Y_0} \right)^\gamma \cdot Y_0.$$

W powyższym zapisie  $Y_{out}$  stanowi wynikową luminancję,  $Y_{in}$  jest oryginalną wartością luminancji danego piksela,  $Y_0$  jest wartością referencyjną, za którą uznano luminancję białego pola tablicy kalibracyjnej zawartej na fotografowanej scenie, a  $\gamma$  jest parametrem stanowiącym o zmianie kontrastu obrazu. Przyjęto, że wartości  $\gamma$  będą wynosiły 0.63, 0.79, 1.00, 1.26 i 1.59. Zbiór wartości został dobrany na podstawie pilotażowego eksperymentu, który pokazał że wartości z wymienionego zakresu są uznawane za najbardziej rzeczywiste. Ostatecznie, by uniknąć zmian koloru wynikających z manipulacji kontrastem, posłużono się podejściem do korekcji koloru opisanym w publikacji [MMTH09].

## Uczestnicy

W badaniu wzięło udział 28 uczestników, w tym 8 kobiet i 20 mężczyzn. Średnia wieku wynosiła 24.3 lat. Ośmioro z nich podczas badania korzystało z okularów korekcyjnych. Wszyscy deklarowali poprawne widzenie stereoskopowe. Żadna z osób nie była poinformowana o celu prowadzonego badania.

## Procedura badawcza

Biorący udział w badaniu mieli za zadanie oglądając pary następujących po sobie stereoskopowych obrazów  $HDR$ , wybierając ten, który uważają za wyglądający

bardziej realistycznie. Obrazy w parach różniły się kontrastem (5 poziomów) i dystansem pomiędzy osiami widzenia obojga oczu (*IAD*, 4 poziomy).

Aby zmniejszyć liczbę testowanych przypadków, a tym samym skrócić czas trwania eksperymentu do około 30 minut, porównywano wyłącznie sąsiadujące ze sobą wartości parametrów. W ten sposób zredukowano liczbę porównań do 55, przy czym każde z nich było powtarzane pięciokrotnie w trakcie trwania sesji eksperymentalnej. Wykorzystując zasadę skali Thurstone'a, na podstawie tak zredukowanej liczby porównań uzupełniono brakujące oceny [SF01a]. Kolejność przedstawianych przypadków była losowa.

### 3.2.2.2 Wyniki

Dzięki zastosowaniu skali Thurstone'a, możliwe było wprowadzenie jednostki minimalnej zmiany pozwalającej na rozróżnienie realizmu (ang. *Just-Noticeable Difference*, *JND*).

#### Klasteryzacja wyników

Próbując na potrzeby analizy uśrednić uzyskane dane zaobserwowano, że odpowiedzi badanych wykazują duże zróżnicowanie. Uznano więc, że nieodpowiednie byłoby rozpatrywanie jedynie średnich wartości. Zdecydowano o podziale rezultatów na grupy charakteryzujące preferencje uczestników badania.

W celu przeprowadzenia klasteryzacji, dokonano regresji parabolicznej wyników oceny realizmu dla każdego z uczestników badania, osobno w uzależnieniu od kontrastu  $\gamma$  i wartości *IAD*. Za parametry klasteryzacji uznano wartość, dla której osiągnęty jest wierzchołek paraboli, a także kąt nachylenia jej ramion dany jej współczynnikiem kwadratowym. Za trzeci parametr klasteryzacji uznano współczynnik kierunkowy prostej po dokonaniu regresji liniowej.

#### Podział według preferencji odnośnie intensywności wskazówki stereoskopowej

Wyróżniono cztery charakterystyczne grupy uczestników pod względem ich preferencji intensywności wskazówki stereoskopowej w kontekście realizmu oglądanego obrazu (zob. Rys. 3.25):

- Preferowana mała stereoskopia lub jej brak (6 osób)
- Preferowana umiarkowana stereoskopia (6 osób)
- Preferowana wzmocniona stereoskopia (8 osób)
- Brak czułości na stereoskopię (8 osób)

Zaskakującą obserwacją jest wskazanie znacznej grupy osób, które nie utożsamiały występowania stereoskopowej wskazówki głębi z realizmem (ostatnia grupa, 8 osób), lub wręcz preferowały jej brak (pierwsza z grup, 6 osób). Zdaje się to być przeciwnym zjawiskiem do trójwymiarowości rozumianej powszechnie jako element

naturalności oglądanej sceny [KKL<sup>+</sup>08]. Z drugiej strony, istniała grupa osób, która za najbardziej realistyczny uznała obraz, w którym wskazówka stereoskopowa została wzmocniona, a więc obraz zawierający rozsunięcie znacznie większe od przeciętnego rozstawu ludzkich oczu. Co ciekawe, osoby które preferowały największe wartości *IAD*, częściej były też entuzjastami wzmocnionego kontrastu. Natomiast osoby bardziej powściągliwe w ocenie realności większych *IAD* przejawiały podobne umiarkowanie w ocenie kontrastu.

### Podział według preferencji kontrastu

Wyróżniono trzy charakterystyczne grupy uczestników pod względem ich preferencji kontrastu w kontekście realizmu oglądanego obrazu (zob. Rys. 3.26):

- Preferowany zmniejszony kontrast (3 osoby)
- Preferowany neutralny kontrast (14 osób)
- Preferowany wzmocniony kontrast (11 osób)

W przeciwieństwie do klasteryzacji z uwagi na *IAD*, nie udało się znaleźć osób które nie byłyby czułe na kontrast. Większe nachylenie krzywych wskazujących preferencje obserwatorów odnośnie kontrastu sugerują, że ta cecha znacznie bardziej wpływała w ich ocenie na wrażenie realizmu.

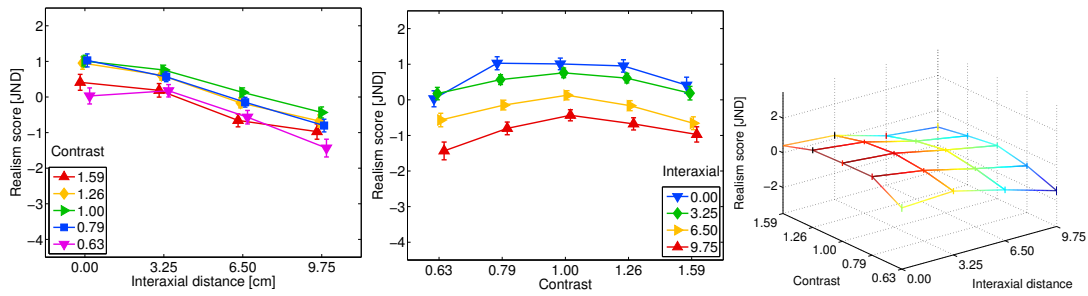
#### 3.2.2.3 Wnioski

Ważnym spostrzeżeniem wynikającym z analizy wyników 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 w opisanych wyżej grupach jest stosunkowo duża. Uzasadnione więc wydaje się wnioskowanie, że każda z nich reprezentuje istotną liczbę osób stanowiących część populacji.

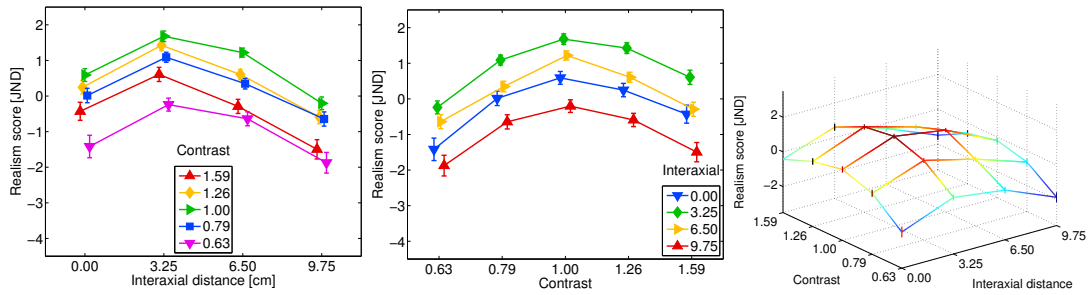
W kwestii preferencji dotyczących kontrastu, można zauważyć trend prowadzący do częstszego wyboru wzmocnionego, niż osłabionego kontrastu jako tego bardziej realnego. Takie zjawisko zostało opisane już wcześniej [YMMS06], gdzie badani wybierali mocniejszy od neutralnego kontrast nawet w przypadku gdy 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.

Warto zauważyć także, że nie obserwuje się ani wzajemnego wpływu zmieniających się wartości *IAD* na preferencje kontrastu, ani odwrotnego. Krzywe na wykresach prezentowanych na Rys. 3.25 i Rys. 3.26 rzadko kiedy się przecinają, zdają się przebiegać niemal w równych odstępach od siebie na całej rozciągłości dziedziny. Można wnioskować, że oba te atrybuty wpływają na postrzeganie realizmu w sposób niezależny.

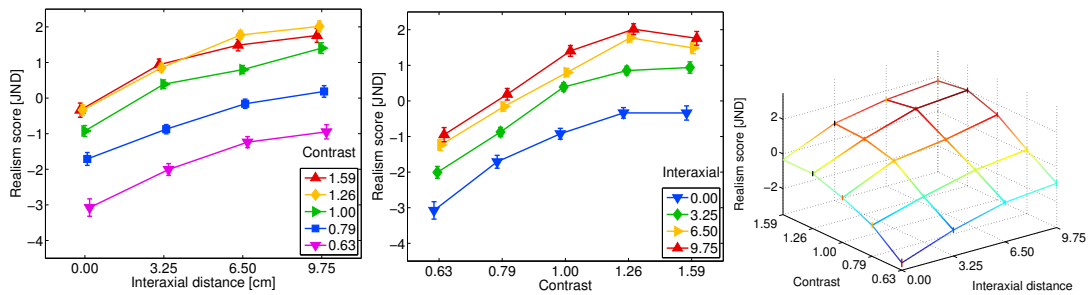
(a) Preferowana mała stereoskopia lub jej brak (6 osób):



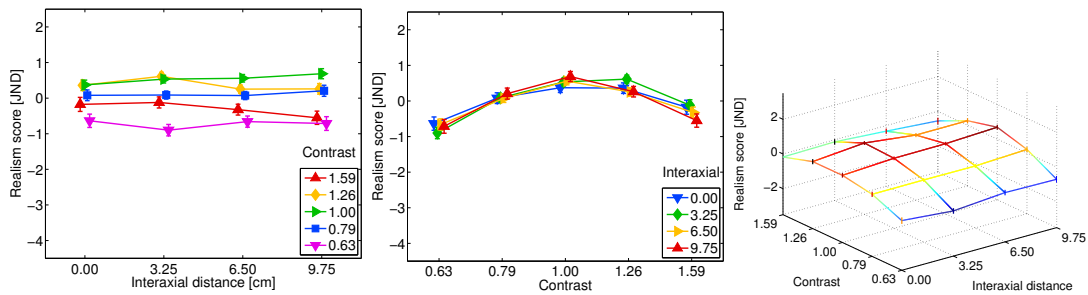
(b) Preferowana umiarkowana stereoskopia (6 osób):



(c) Preferowana wzmocniona stereoskopia (8 osób):

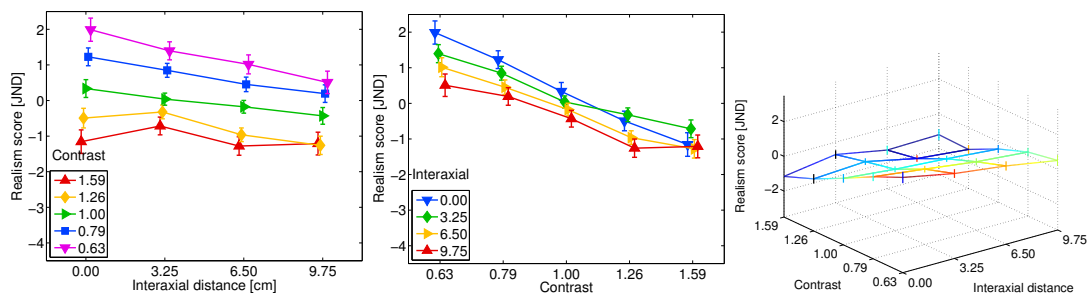


(d) Brak czułości na stereoskopię (8 osób):

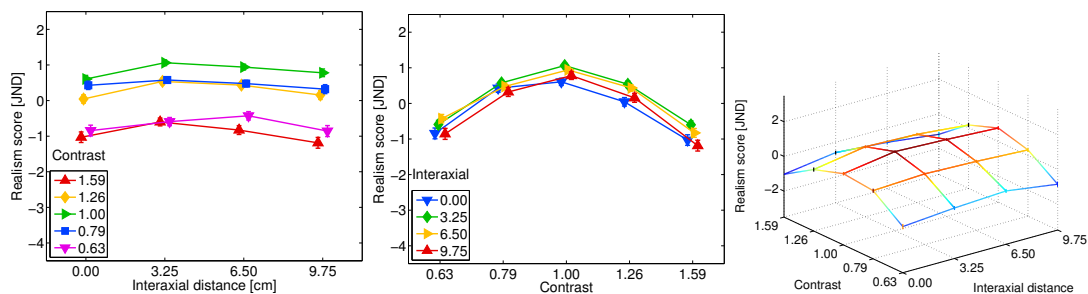


Rysunek 3.25: Rezultaty badania przedstawione po klasteryzacji ze względu na wartości  $IAD$ . Kolejno od lewej: wrażenie realizmu w uzależnieniu od wartości  $IAD$  (kryterium klasteryzacji), wrażenie realizmu w uzależnieniu od wartości  $\gamma$  kontrastu, obie zależności przedstawione jako powierzchnia.

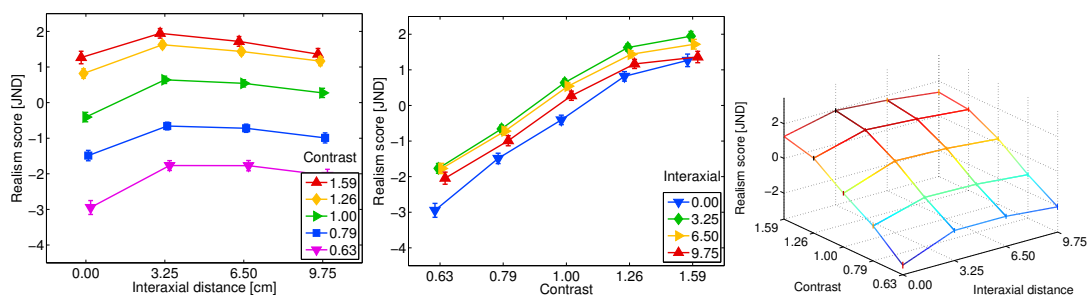
(a) Preferowany zredukowany kontrast (3 osoby):



(b) Preferowany neutralny kontrast (14 osób):



(c) Preferowany wzmocniony kontrast (11 osób):



Rysunek 3.26: Rezultaty badania przedstawione po klasteryzacji ze względu na wartości  $\gamma$  kontrastu. Kolejno od lewej: wrażenie realizmu w uzależnieniu od wartości  $IAD$  (kryterium klasteryzacji), wrażenie realizmu w uzależnieniu od wartości  $\gamma$  kontrastu, obie zależności przedstawione jako powierzchnia.

## Podsumowanie rozdziału

W rozdziale zaproponowano technikę interaktywnej symulacji zjawiska głębi ostrości, która pozwala na wykorzystanie monokularnej wskazówki głębi jaką jest rozmycie obrazu na powierzchni siatkówki, bez poświęcania w zamian zdolności obserwatora do oglądania obrazu w sposób przez niego wybrany i w wybranej przez niego kolejności. Pokazano sposób implementacji tej techniki oparty o wykorzystanie programowalnego procesora graficznego. Przedstawione rozwiązanie zostało poddane ewaluacji w celu sprawdzenia jego skuteczności. Badane osoby preferowały symulację głębi ostrości działającą w sposób interaktywny, opartą o zbierane na bieżąco dane okulograficzne. To rozwiązanie zostało także wskazane jako wpływające pozytywnie na realizm oglądanego obrazu.

Pokazano, że interaktywna symulacja głębi ostrości może także służyć złagodzeniu skutków konfliktu konwergencji i akomodacji, uznawanego za podstawową przyczynę dyskomfortu podczas korzystania z wyświetlaczy stereoskopowych. W ten sposób możliwe jest ograniczenie zmęczenia wzroku i bólu głowy, uzupełniając brakujący element w postaci rozmycia obrazu, które powinno być skorelowane z ruchami konwergencyjnymi oczu.

Zweryfikowano hipotezę o wpływie kontrastu na postrzeganie głębi w wyświetlaczu stereoskopowym, wykazując eksperymentalnie że w warunkach zbliżonych do progu rozróżnialności odległości na podstawie wskazówek binokularnych, duża różnica kontrastu jest w stanie przelożyć się na zmianę postrzeganego dystansu.

Wykazano również, że kontrast jest istotnym czynnikiem wpływającym na postrzeganie wyświetlanych obrazów jako realistycznych. Obserwacja, która do tej pory znana była dla wyświetlaczy monoskopowych, została potwierdzona także dla warunków stereoskopowych i obrazów o poszerzonym zakresie dynamiki luminancji. Jest to o tyle istotne, że współczesne wyświetlacze stereoskopowe bazują najczęściej na technice związanej z okularami migawkowymi, co prowadzi do znaczącego pogorszenia kontrastu wyświetlanego obrazu poprzez skrócenie czasu ekspozycji każdej z klatek przeznaczonych dla pojedynczego oka. Należy mieć na uwadze to, że w takiej sytuacji wrażenia realności oglądanego obrazu u użytkowników mogą być znacznie mniejsze.

# 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 obserwacji wizualizacji wirtualnej sceny. Na łamach niniejszej 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.

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 tym samym znane wcześniej zjawisko, obserwowane do tej pory wyłącznie w przypadku wyświetlaczy o wąskim zakresie dynamiki. Dokonano określenia związku pomiędzy występowaniem obrazowania zjawiska głębi ostrości, a wrażeniem głębi i zanurzenia w trójwymiarowej scenie. Opisano także wpływ kontrastu dla obrazów stereoskopowych w warunkach poszerzonego zakresu dynamiki luminancji na poczucie realności u obserwatora.

Stosowana część badań zawartych w pracy, dotyczyła opracowania sposobu realizacji interaktywnej symulacji zjawiska głębi ostrości. Badano też możliwość użycia proponowanej techniki do złagodzenia konfliktu konwergencji i akomodacji w przypadku wyświetlania obrazów stereoskopowych.

Wykorzystanie technik okulograficznych w czasie rzeczywistym do zmiany cech prezentowanego obrazu jest podejściem nowatorskim, opisanym w literaturze w bardzo ograniczonym zakresie. Podczas prac nad nimi, koniecznością okazało się przezwyciężenie istotnych problemów związanych z dokładnością identyfikacji elementów obrazu na których spoczywa uwaga obserwatora. W tym celu opracowano nową metodę identyfikacji obiektu uwagi nazwaną *GDOT*, która przewyższa swoją skutecznością dotychczas stosowane techniki we wspomnianych warunkach. Metoda ta może być wykorzystana także w innych obszarach zastosowań, czego przykładem jest, zaproponowane i przebadane w ramach pracy, zagadnienie analizy uwagi osoby oglądającej obraz wideo.



## Wnioski końcowe

Na podstawie opisanych w rozprawie prac badawczych, wyciągnięto następujące wnioski:

- Probabilistyczna metoda oparta zarówno o porównanie położeń, jak i wektorów prędkości punktu skupienia wzroku oraz wyodrębnionych elementów prezentowanego obrazu, umożliwia efektywną identyfikację obiektu będącego przedmiotem uwagi obserwatora (Rozdział 2.3, opublikowano w [MBM13]).
- Zaproponowana w pracy probabilistyczna metoda identyfikacji przedmiotu uwagi *GDOT* jest odpowiednia dla analizy uwagi wzrokowej osoby oglądającej materiał wideo, gdzie zbiór atraktorów uwagi powstaje z pomocą metody estymacji przepływu optycznego na obrazie (Rozdział 2.4.1, opublikowano w [BM14a]).
- Wiedza o chwilowym punkcie skupienia wzroku użytkownika może być wykorzystana do przyspieszenia syntezy obrazu, poprzez redukcję jakości obrazu w obszarach niepokrytych widzeniem centralnym (Rozdział 2.4.2, opublikowano w [TB10]).
- Wprowadzenie elementów interfejsu użytkownika opartych na wiedzy o chwilowym punkcie skupienia wzroku użytkownika, przyczynia się do zwiększenia poczucia zanurzenia u użytkownika systemu wirtualnej rzeczywistości (Rozdział 2.4.3, opublikowano w [BM14b]).
- Interaktywna, oparta na wiedzy o punkcie skupienia wzroku obserwatora symulacja zjawiska głębi ostrości podczas wizualizacji wirtualnego środowiska, przyczynia się do zwiększenia poczucia realizmu oglądanego obrazu (Rozdział 3.1, opublikowano w [Baz10, MBT11]).
- Wykorzystanie interaktywnej symulacji głębi ostrości przy obrazowaniu stereoskopowym, w połączeniu z wiedzą o odległości na której użytkownik skupia wzrok, pozwala złagodzić skutki konfliktu konwergencji-akomodacji (Rozdział 3.1.4, opublikowano w [Duc14]).
- Kontrast wpływa na postrzeganie odległości w obrazie wyświetlanym na ekranie o poszerzonym zakresie dynamiki luminancji, jednak wpływ ten jest niewielki w sytuacji gdy obraz dostarcza także silnej wskazówki stereoskopowej (Rozdział 3.2.1, opublikowano w [Pet14]).
- W populacji występują charakterystyczne grupy cechujące się silnymi preferencjami odnośnie kontrastu i siły stereoskopowej wskazówki głębi, objawiające się skorelowanym z nimi poczuciem realizmu oglądanego obrazu (Rozdział 3.2.2, opublikowano w [Pet14]).

Powyższe wnioski wskazują na zawarte w pracy elementy nowości dla poruszanego obszaru nauki.

## Odniesienie do tezy

Na początku rozprawy postawiono 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.*

W toku prac, poprzez eksplorację możliwości wpłynięcia na postrzeganie głębi za pomocą elementów dostępnych w graficznych systemach komputerowych, pokazano i zbadano techniki pozwalające na poprawę odbioru informacji o przestrzenności wirtualnego środowiska. Tym samym przyczyniono się do poprawy jakości wizualizacji tychże, rozumianej jako zdolność użytkownika systemu komputerowego do wykształcenia poczucia zanurzenia w symulowanej, trójwymiarowej rzeczywistości.

Posłużono się wiedzą o chwilowym kierunku spojrzenia użytkownika, by imitować działanie procesu akomodacji. Technika ta przyczyniła się zarówno do wzbogacenia obrazu o dodatkową wskazówkę głębi, jak i do złagodzenia problemu współczesnych wyświetlaczy stereoskopowych. Zamodelowano powiązanie kontrastu na wyświetlaczu o poszerzonym zakresie dynamiki luminancji z wrażeniem odległości oglądanych obiektów, a także zbadano powiązanie pomiędzy kontrastem i rozbieżnością dwuoczną a poczuciem realizmu obrazu.

W ramach rozprawy potwierdzono zatem postawioną wcześniej tezę.

## Dalsze prace

Niniejsza praca skupiła się jedynie na wybranych elementach ludzkiego mechanizmu postrzegania obrazu. Dążenie do uzyskania jak najwierniejszego odwzorowania realnych bodźców za pomocą graficznych systemów komputerowych, wymaga dalszych badań nad możliwościami wpływu na percepcję człowieka.

Udało się zamodelować związek pomiędzy wskazówką głębi pochodzącą z kontrastu a widzeniem stereoskopowym, przy wywoływaniu wrażenia przestrzenności obrazu. Efektywne wykorzystanie dostępnych możliwości komunikowania głębi i realizmu obrazu wymaga, by rozpatrzyć i zbadać także inne związki pomiędzy poszczególnymi wskazówkami.

Proponowana w rozprawie interaktywna symulacja głębi ostrości stanowi rozwiązanie, które mogłoby być zastosowane dla wyświetlaczy oferujących wiele płaszczyzn akomodacji. Zbadanie w takich warunkach skutków występowania rozmycia obrazu zależnego od kierunku spojrzenia, przy jednoczesnym zastosowaniu wyświetlania stereoskopowego, mogłoby przyczynić się do ostatecznego potwierdzenia genezy dyskomfortu podczas korzystania z tego rodzaju wyświetlaczy i znalezienia sposobów na dalszą jego redukcję. Technika ta wydaje się również być ciekawym rozwiązaniem dla systemów wirtualnej rzeczywistości opartych o wyświetlacze *HMD* (ang. *Head-Mounted Display*), a także systemy wykorzystujące matryce mikrosoczewek o zmiennych, sterowalnych cechach optycznych.

Intrygującym obszarem zainteresowania zdają się być także inne sposoby wykorzystania okulografii i technik identyfikacji przedmiotu uwagi wzrokowej użytkownika wirtualnej rzeczywistości, a także dalsza poprawa jakości działania samych metod. Uniwersalność zaproponowanej metody *GDOT* skutkuje szerokim spektrum potencjalnych zastosowań obejmujących nie tylko syntezę obrazu. Ciekawym kierunkiem rozwoju może być wykorzystanie tej metody dla zredukowania uciążliwości procesu kalibracji okulografu lub całkowitego wyeliminowania tejże. To ostatnie mogłoby okazać się możliwe poprzez opracowanie techniki autokalibracyjnej, bazującej na zbieranych na bieżąco danych o oglądaniu dynamicznej sceny.

Z kolei użycie binokularnego okulografu w kontekście szacowania odległości, na której użytkownik skupia wzrok, pozwala na szerokie zastosowanie pozyskanej w ten sposób informacji. Poza wykorzystaniem jej do symulacji zjawiska głębi ostrości, możliwe byłoby także posłużenie się tą wiedzą dla poprawy skuteczności metody *GDOT* poprzez dodanie trzeciego wymiaru dla jej danych wejściowych.

Autor w swojej przyszłej pracy naukowej chciałby zgłębić te właśnie obszary, przyczyniając się do zwiększania możliwości współczesnych systemów wizualizacji w zakresie przekonującej symulacji wirtualnych środowisk.

# 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.
- [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.

# Literatura

- [AEC96] R.S. Allison, E. Eizenman, and B.S.K. Cheung. Combined head and eye tracking system for dynamic testing of the vestibular system. *Biomedical Engineering, IEEE Transactions on*, 43(11):1073–1082, Nov 1996.
- [APP<sup>+</sup>09] Zainal Arief, Djoko Purwanto, Dadet Pramadihanto, Tetsuo Sato, and Kotaro Minato. Relation between eye movement and fatigue: Classification of morning and afternoon measurement based on fuzzy rule. In *Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME), 2009 International Conference on*, pages 1–6. IEEE, 2009.
- [BC04] Emily Brown and Paul Cairns. A grounded investigation of game immersion. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1297–1300, New York, NY, USA, 2004. ACM.
- [BF85] W. Becker and A. F. Fuchs. Prediction in the oculomotor system: smooth pursuit during transient disappearance of a visual target. *Experimental Brain Research*, 57.3:562–575, 1985.
- [BK08] Brian A. Barsky and Todd J. Kosloff. Algorithms for rendering depth of field effects in computer graphics. In *Proceedings of the 12th WSEAS International Conference on Computers*, ICCOMP'08, pages 999–1010. World Scientific and Engineering Academy and Society (WSEAS), 2008.
- [Bli09] Pieter Bignaut. Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics*, 71(4):881–895, 2009.
- [BM83] A. Terry Bahill and Jack D. McDonald. Model emulates human smooth pursuit system producing zero-latency target tracking. *Biological Cybernetics*, pages 48–213, 1983.
- [BM07] Doug A Bowman and Ryan P McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7), 2007.

- [BRG81] R. Buurman, T. Roersma, and J.F. Gerrissen. Eye movements and the perceptual span in reading. *Reading Research Quarterly*, pages 227–235, 1981.
- [Bri07] Robert Bridson. Fast poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH 2007 Sketches*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.
- [Bro58] D. E. Broadbent. *Perception and Communication*. Pergamon Press, 1958.
- [BTHC03] Brian A. Barsky, Michael J. Tobias, Daniel R. Horn, and Derrick P. Chu. Investigating occlusion and discretization problems in image space blurring techniques. In *International Conference on Vision, Video, and Graphics*, pages 97–102, 2003.
- [CAS<sup>+</sup>91] Christine A. Curcio, Kimberly A. Allen, Kenneth R. Sloan, Connie L. Lerea, James B. Hurley, Ingrid B. Klock, and Ann H. Milam. Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *The Journal of Comparative Neurology*, 312(4):610–624, 1991.
- [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.
- [CG87] J. R. Carl and R. S. Gellman. Human smooth pursuit: stimulus-dependent responses. *Journal of Neurophysiology*, 57(5):1446–1463, 1987.
- [CM16] Michał Chwesiuk and Radosław Mantiuk. Acceptable System Latency for Gaze-Dependent Level of Detail Rendering. In Luis Gonzaga Magalhaes and Rafal Mantiuk, editors, *EG 2016 - Posters*. The Eurographics Association, 2016.
- [Cor56] Tom N Cornsweet. Determination of the stimuli for involuntary drifts and saccadic eye movements. *JOSA*, 46(11):987–988, 1956.
- [CPC84] Robert L. Cook, Thomas Porter, and Loren Carpenter. Distributed ray tracing. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '84, pages 137–145, New York, NY, USA, 1984. ACM.
- [CW77] WN Charman and Helen Whitefoot. Pupil diameter and the depth-of-field of the human eye as measured by laser speckle. *Journal of Modern Optics*, 24(12):1211–1216, 1977.

- [DC07] Andrew T. Duchowski and ARZU CÖLTEKIN. Foveated gaze-contingent displays for peripheral lod management, 3d visualization, and stereo imaging. *ACM TRANS. MULTIMEDIA COMPUT. COMMUN. APPL. (TOMCCAP). ARCH*, 2007.
- [Dem04] Joe Demers. Depth of field: A survey of techniques. In Randima Fernando, editor, *GPU Gems*, pages 375–390. Addison-Wesley, 2004.
- [DG52] RW Ditchburn and BL Ginsborg. Vision with a stabilized retinal image. *Nature*, 170:36–37, 1952.
- [dG00] Beatrice de Gelder. More to seeing than meets the eye. *Science*, 289(5482):1148–1149, 2000.
- [Dod03] Raymond Dodge. Five types of eye movement in the horizontal meridian plane of the field of regard. *American Journal of Physiology–Legacy Content*, 8(4):307–329, 1903.
- [DPHW11a] 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.
- [DPHW11b] 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.
- [Duc07] Andrew Duchowski. *Eye tracking methodology: Theory and practice*. Springer-Verlag, London, UK, 2007.
- [DY10] Michael Tri Hoang Do and King-Wai Yau. Intrinsically photosensitive retinal ganglion cells. *Physiological Reviews*, 90(4):1547–1581, 2010.
- [EDP<sup>+</sup>12] Kai Essig, Daniel Dornbusch, Daniel Prinzhorn, Helge Ritter, Jonathan Maycock, and Thomas Schack. Automatic analysis of 3d gaze coordinates on scene objects using data from eye-tracking and motion-capture systems. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA ’12, pages 37–44, New York, NY, USA, 2012. ACM.
- [EH08] Jr. Earl Hammon. Practical post-process depth of field. In Hubert Nguyen, editor, *GPU Gems 3*, pages 583–606. NVIDIA Corporation, 2008.
- [EML13] Haider K. Easa, Rafał K. Mantiuk, and Ik Soo Lim. Evaluation of monocular depth cues on a high-dynamic-range display for visualization. *ACM Trans. Appl. Percept.*, 10(3):16, 2013.



- [ES93] Y. Ebisawa and S.-i. Satoh. Effectiveness of pupil area detection technique using two light sources and image difference method. In *Engineering in Medicine and Biology Society, 1993. Proceedings of the 15th Annual International Conference of the IEEE*, pages 1268–1269, 1993.
- [EV95] Casper J Erkelens and Ingrid MLC Vogels. The initial direction and landing position of saccades. *Studies in Visual Information Processing*, 6:133–144, 1995.
- [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.
- [For73] G. David Forney. The viterbi algorithm. In *Proceedings of the IEEE*, pages 268–278, 1973.
- [FPS<sup>+</sup>12] Carsten Finke, Luisa Maria Pech, Carina Sömmmer, Jeremias Schlichting, Sarah Stricker, Matthias Endres, Florian Ostendorf, Christoph J Ploner, Alexander U Brandt, and Friedemann Paul. Dynamics of saccade parameters in multiple sclerosis patients with fatigue. *Journal of neurology*, 259(12):2656–2663, 2012.
- [Fut12] Futuremark. 3DMark 11 White Paper. Technical report, Futuremark, 11 2012.
- [GFD<sup>+</sup>12] Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. Foveated 3d graphics. *ACM Transactions on Graphics (TOG)*, 31(6):164, 2012.
- [GK65] G. Golub and W. Kahan. Calculating the singular values and pseudo-inverse of a matrix. In *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis, Vol. 2, No. 2*, pages 205–224. Society for Industrial and Applied Mathematics, 1965.
- [HA90] Paul Haeberli and Kurt Akeley. The accumulation buffer: Hardware support for high-quality rendering. In *Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '90*, pages 309–318, New York, NY, USA, 1990. ACM.
- [HCB12] Robert T Held, Emily A Cooper, and Martin S Banks. Blur and disparity are complementary cues to depth. *Current biology*, 22(5):426–431, 2012.
- [HE98] Ignace Th.C. Hooge and Casper J. Erkelens. Adjustment of fixation duration in visual search. *Vision Research*, 38(9):1295 – IN4, 1998.
- [Hee06] David Heeger. Perception lecture notes: Depth, size, and shape, 2006.

- [HFC05] J. Hailin, P. Favaro, and R. Cipolla. Visual tracking in the presence of motion blur. *Proc. of Computer Vision and Pattern Recognition (CVPR'05)*, 2:18–25, 2005.
- [HGAB08] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, 2008.
- [HH03] John M Henderson and Andrew Hollingworth. Global transsaccadic change blindness during scene perception. *Psychological Science*, 14(5):493–497, 2003.
- [HNA<sup>+</sup>11a] 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.
- [HNA<sup>+</sup>11b] K. Holmqvist, M. Nystrom, R. Andersson, J. Jarodzka, and J. van de Weijer. *Eye-tracking data and dependent variables*. Oxford University Press, 2011.
- [HR95] Ian P Howard and Brian J Rogers. *Binocular vision and stereopsis*. Oxford University Press, USA, 1995.
- [HR02] I. P. Howard and B. J. Rogers. *Seeing in Depth, volume II: Depth Perception*. University of Toronto Press, 2002.
- [HTA<sup>+</sup>05] L Elliot Hong, Malle Tagamets, Matthew Avila, Ikwunga Wonodi, Henry Holcomb, and Gunvant K Thaker. Specific motion processing pathway deficit during eye tracking in schizophrenia: a performance-matched functional magnetic resonance imaging study. *Biological psychiatry*, 57(7):726–732, 2005.
- [HZ03] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003.
- [IKA07] Shigeru Ichihara, Norimichi Kitagawa, and Hiromi Akutsu. Contrast and depth perception: Effects of texture contrast and area contrast. *Perception*, 36:686–695, 2007.
- [IKT09] Tsuneto Iwasaki, Toshiaki Kubota, and Akihiko Tawara. The tolerance range of binocular disparity on a 3d display based on the physiological characteristics of ocular accommodation. *Displays*, 30(1):44–48, 2009.
- [IM06] Poika Isokoski and Benoit Martin. Eye tracker input in first person shooter games. In *Proceedings of the 2nd Conference on Communication by Gaze Interaction*, pages 76–79. Communication by Gaze Interaction (COGAIN), 2006.

- [IRA06] Victoria Interrante, Brian Ries, and Lee Anderson. Distance perception in immersive virtual environments, revisited. In *Virtual Reality Conference, 2006*, pages 3–10. IEEE, 2006.
- [itu12] ITU Gaze Tracker software. <http://www.gazegroup.org/home>, 2012. Dostęp: 2012-08-20.
- [Jac91] Robert J. K. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Transactions on Information Systems*, 9:152–169, 1991.
- [Jav79] Louis Emile Javal. Essai sur la physiologie de la lecture. In *Annales D’Oculistique*, 1879.
- [JB97] TJWM Janssen and FJJ Blommaert. Image quality semantics. *Journal of Imaging Science and Technology*, 41(5):555–560, 1997.
- [JJ05] Erika Jönsson and Erika Jönsson. If looks could kill - an evaluation of eye tracking in computer games. Technical report, 2005.
- [Kar00] Keith S Karn. “saccade pickers” vs. “fixation pickers”: the effect of eye tracking instrumentation on research. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 87–88. ACM, 2000.
- [Kaw03] Masaki Kawase. Frame Buffer Postprocessing Effects in DOUBLE-S.T.E.A.L (Wreckless). Technical report, Bunkasha Games, 2003.
- [Kes16] John Kessenich. The opengl shading language 4.50. Technical report, The Khronos Group Inc., 2016.
- [KKL<sup>+</sup>08] R. G. Kaptein, A. Kuijsters, M. T. M. Lambooij, W. A. IJsselsteijn, and I. Heynderickx. Performance evaluation of 3D-TV systems. In *Proc. Electron. Imaging Image Quality Syst. Perform. V*, volume 6808, pages 19:1–11. SPIE, 2008.
- [KMH95] Craig Kolb, Don Mitchell, and Pat Hanrahan. A realistic camera model for computer graphics. In *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ’95*, pages 317–324, New York, NY, USA, 1995. ACM.
- [KvI<sup>+</sup>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.
- [Lea13] Anna Leach. Samsung pushes eye-tracking in Galaxy S4. <http://blogs.wsj.com/tech-europe/2013/03/15/samsung-pushes-eye-tracking-in-galaxy-s4/>, 2013. Dostęp: 2014-08-20.

- [LES09] Sungkil Lee, Elmar Eisemann, and Hans-Peter Seidel. Depth-of-field rendering with multiview synthesis. In *ACM SIGGRAPH Asia 2009 Papers*, SIGGRAPH Asia '09, pages 134:1–134:6, New York, NY, USA, 2009. ACM.
- [LFH09] Marc Lambooi, Marten Fortuin, and Ingrid Heynderickx. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53:1–14, 2009.
- [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.
- [LM04] J. Leyba and J. Malcolm. Eye tracking as an aiming device in a computer game, 2004.
- [LRSK13] Sophie N Lanthier, Evan F Risko, Daniel Smilek, and Alan Kingstone. Measuring the separate effects of practice and fatigue on eye movements during visual search. In *Proceedings of COGSCI 2013*. Cognitive Science Society, 2013.
- [LWP05] Dongheng Li, David Winfield, and Derrick J Parkhurst. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 79–79. IEEE, 2005.
- [LZ99] R John Leigh and David S Zee. *The neurology of eye movements*, volume 90. Oxford University Press New York, 1999.
- [Mar82] D Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Co., 1982.
- [MB98] Brad C. Motter and Eric J. Belky. The zone of focal attention during active visual search. *Vision Research*, 38(7):1007 – 1022, 1998.
- [MCMH04] Susana Martinez-Conde, Stephen L Macknik, and David H Hubel. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3):229–240, 2004.
- [MD11] Martin Mittring and Bryan Dudash. The Technology Behind the DirectX 11 Unreal Engine "Samaritan" Demo. Technical report, Epic Games, NVIDIA, 2011.
- [Mil14] M. Millodot. *Dictionary of Optometry and Visual Science*. Elsevier Health Sciences UK, 2014.

- [MJ12] Radosław Mantiuk and Sebastian Janus. Gaze-dependent ambient occlusion. *Lecture Notes in Computer Science (Proc. of ISVC 2012)*, 7431:523–532, 2012.
- [MKMS07] Rafał Mantiuk, Grzegorz Krawczyk, Radosław Mantiuk, and Hans-Peter Seidel. High dynamic range imaging pipeline: Perception-motivated representation of visual content. In Bernice E. Rogowitz, Thrasyvoulos N. Pappas, and Scott J. Daly, editors, *Proc. Human Vision and Electronic Imaging XII*, volume 6492. SPIE, 2007.
- [MM05] Carlos H Morimoto and Marcio RM Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.
- [MMN99] Susana Marcos, Esther Moreno, and Rafael Navarro. The depth-of-field of the human eye from objective and subjective measurements. *Vision research*, 39(12):2039–2049, 1999.
- [MMTH09] Radosław Mantiuk, Rafał Mantiuk, Anna Tomaszewska, and Wolfgang Heidrich. Color correction for tone mapping. *Computer Graphics Forum*, 28(2):193–202, 2009.
- [MN84] Suzanne P McKee and Ken Nakayama. The detection of motion in the peripheral visual field. *Vision Res.*, 24(1):25–32, 1984.
- [Mon00] J. Monaco. *How to Read a Film: The World of Movies, Media, and Multimedia: Language, History, Theory*. Oxford University Press, 3rd edition, 2000.
- [MRD12] L. McIntosh, B. E. Riecke, and S. DiPaola. Efficiently simulating the bokeh of polygonal apertures in a post-process depth of field shader. *Computer Graphics Forum*, 31(6):1810–1822, 2012.
- [NS71] David Noton and Lawrence Stark. Eye movements and visual perception. *Scientific American*, 1971.
- [OBO94] Robert P O’Shea, Shane G Blackburn, and Hiroshi Ono. Contrast as a depth cue. *Vision Research*, 34(12):1595–1604, 1994.
- [oNIRP<sup>+</sup>13] International Commission on Non-Ionizing Radiation Protection et al. Icnirp guidelines on limits of exposure to incoherent visible and infrared radiation. *Health Physics*, 105(1):74–96, 2013.
- [Ost35] G. Osterberg. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica. Supplement*, 6:1–103, 1935.
- [PC82] Michael Potmesil and Indranil Chakravarty. Synthetic image generation with a lens and aperture camera model. *ACM Trans. Graph.*, 1(2):85–108, April 1982.

- [PCS<sup>+</sup>00] Emilee Patrick, Dennis Cosgrove, Aleksandra Slavkovic, Jennifer A Rode, Thom Verratti, and Greg Chiselko. Using a large projection screen as an alternative to head-mounted displays for virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 478–485. ACM, 2000.
- [Pfe12] Thies Pfeiffer. Measuring and visualizing attention in space with 3d attention volumes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 29–36, New York, NY, USA, 2012. ACM.
- [PKS<sup>+</sup>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.
- [Pog12] Tomaso Poggio. The levels of understanding framework, revised. *Perception*, 41(9):1017–1023, 2012.
- [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.
- [PWBoNR73] J.F. Parker, V.R. West, Inc Biotechnology, and United States. Office of Naval Research. *Bioastronautics Data Book*. NASA SP. Scientific and Technical Information Office, National Aeronautics and Space Administration, 1973.
- [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.
- [RHFL10] Stephan Reichelt, Ralf Häussler, Gerald Fütterer, and Norbert Lester. Depth cues in human visual perception and their realization in 3d displays. volume 7690, pages 76900B–76900B–12, 2010.
- [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.
- [Rok96] Przemysław Rokita. Generating depth of-field effects in virtual reality applications. *IEEE Computer Graphics and Applications*, 16(2):18–21, Mar 1996.
- [Rok97] Przemysław Rokita. Application of image processing techniques in computer graphics algorithms. *Computer Networks and ISDN Systems*, 29(14):1705 – 1714, 1997.

- [Rol53] Wilhelm Rollmann. Zwei neue stereoskopische methoden. *Annalen der Physik*, 166(9):186–187, 1853.
- [RWH<sup>+</sup>16] T. Roth, M. Weier, A. Hinkenjann, Y. Li, and P. Slusallek. An analysis of eye-tracking data in foveated ray tracing. In *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)*, pages 69–73, Oct 2016.
- [SAMHV09] J. C. San Agustin, J. Mateo, J.P. Hansen, and A. Villanueva. Evaluation of the potential of gaze input for game interaction. *PsychNology Journal*, 7:213–236, 2009.
- [Sco92] Cary Scofield. Graphics gems III. chapter 2 1/2-D Depth-of-field Simulation for Computer Animation, pages 36–38. Academic Press Professional, Inc., San Diego, CA, USA, 1992.
- [SE80] T.A. Salthouse and C.L. Ellis. Determinants of eye-fixation duration. *American Journal of Psychology*, 93(2):207–234, 1980.
- [Sen09] SensoMotoric Instruments GmbH. *RED250 Technical Specification*, 2009.
- [Sen10] SensoMotoric Instruments GmbH. *iViewX Manual*, 2010.
- [SF01a] D. Silverstein and J. Farrell. Efficient method for paired comparison. *Journal of Electronic Imaging*, 10:394–398, 2001.
- [SF01b] DA Silverstein and JE Farrell. Efficient method for paired comparison. *Journal of Electronic Imaging*, 10:394, 2001.
- [SFHHFLL91] D. Scott, J.M. Findlay, Winchester Hursley Human Factors Laboratory, and IBM UK Hursley Human Factors Laboratory. *Visual Search, Eye Movements and Display Units*. HF–147. IBM UK Hursley Human Factors Laboratory, 1991.
- [SFW<sup>+</sup>11] V. De Silva, A. Fernando, S. Worrall, H. Kodikara Arachchi, and A. Kondo. Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-d displays. *IEEE Transactions on Multimedia*, 13(3):498–506, June 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.
- [SG06] J. David Smith and T. C. Nicholas Graham. Use of eye movements for video game control. In *In Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (p. 20)*. ACM Press, 2006.

- [Sie14] Adam Siekawa. Gaze-dependent ray tracing. In *Proceedings of Central European Seminar on Computer Graphics 2014*, Vienna, Austria, 2014. VUT.
- [SKB<sup>+</sup>05] Miriam Spering, Dirk Kerzel, Doris I Braun, Michael J Hawken, and Karl R Gegenfurtner. Effects of contrast on smooth pursuit eye movements. *Journal of Vision*, 5(5):6, 2005.
- [Sla03] Mel Slater. A note on presence terminology. *Presence connect*, 3(3):1–5, 2003.
- [Sla09] Mel Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.
- [Sou13] Tiago Sousa. CryEngine 3 Graphics Gems. Technical report, Crytek, 2013.
- [ST04] Thorsten Scheuermann and Natalya Tatarchuk. Advanced depth of field rendering. *ShaderX3: Advanced Rendering with DirectX and OpenGL*, 2004.
- [Str14] Filip Strugar. An investigation of fast real-time gpu-based image blur algorithms. Technical report, Intel, 7 2014.
- [SW97] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments five: Speculations on the role of presence in virtual environments. *Presence: Teleoper. Virtual Environ.*, 6(6):603–616, December 1997.
- [SW12] David Schedl and Michael Wimmer. A layered depth-of-field method for solving partial occlusion. *Journal of WSCG*, 20(3):239–246, June 2012.
- [Sze10] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010.
- [TC75] Jill Tucker and WN Charman. The depth-of-focus of the human eye for snellen letters. *Optometry & Vision Science*, 52(1):3–21, 1975.
- [TG80] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.
- [TIK90] Akira Tomono, Muneo Iida, and Yukio Kobayashi. A tv camera system which extracts feature points for non-contact eye movement detection. In *1989 Advances in Intelligent Robotics Systems Conference*, pages 2–20. International Society for Optics and Photonics, 1990.



- [TML<sup>+</sup>04] Yukari Takarae, Nancy J Minshew, Beatriz Luna, Christine M Krisky, and John A Sweeney. Pursuit eye movement deficits in autism. *Brain*, 127(12):2584–2594, 2004.
- [Tob09] Tobii Technology AB. *Tobii T/X series Eye Trackers. Product Description*, 2.0 edition, 2009.
- [tob11] Accuracy and precision test method for remote eyetrackers. Technical report, Tobii Technology, 2011.
- [tob13] Tobii mobile device stand for x2. <http://www.tobii.com/en/eye-tracking-research/global/products/hardware-accessories/mobile-device-stand-X2/>, 2013. Do-step: 2014-08-20.
- [TSB11] D. M. Hoffman T. Shibata, J. Kim and M. S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8):1 – 29, 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.
- [VGS13] Matteo Valsecchi, Karl R Gegenfurtner, and Alexander C Schütz. Saccadic and smooth-pursuit eye movements during reading of drifting texts. *Journal of vision*, 13(10):8, 2013.
- [Voß09] Adrian Voßkühler. Ogama description (for version 2.5). a software to record, analyze and visualize gaze and mouse movements in screen based environments., 2009.
- [Wat88] R. J. Watt. *Visual processing: computational, psychophysical and cognitive research*. Lawrence Erlbaum, 1988.
- [WBB11] John White and Coliln Barre-Brisebois. More Performance! Five Rendering Ideas from Battlefield 3 and Need For Speed: The Run. Technical report, Black Box, DICE, Electronic Arts, 2011.
- [WCF89] Jeremy M Wolfe, Kyle R Cave, and Susan L Franzel. Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and performance*, 15(3):419, 1989.
- [Whe38] Charles Wheatstone. Contributions to the physiology of vision.–Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128:371–394, 1838.
- [Wid84] Heino Widdel. Operational problems in analysing eye movements. *Advances in Psychology*, 22:21–29, 1984.

- [Wil06] Lippincott Williams & Wilkins, editor. *Medical Dictionary*. Med-Lexicon International Ltd., Bexhill-on-Sea, UK, 2006.
- [WK81] Heino Widdel and Jürgen Kaster. Eye movement measurement in the assessment and training of visual performance. In J. Moraal and K.-F. Kraiss, editors, *Manned Systems Design*, volume 17 of *NATO Conference Series*, pages 251–270. Springer US, 1981.
- [Wol07] Jeremy M Wolfe. Guided search 4.0. pages 99–119. Oxford University Press New York, NY, 2007.
- [Wol13] D. Wolff. *OpenGL 4 Shading Language Cookbook, Second Edition*. Packt Publishing, 2013.
- [Wol15] Krzysztof Wolski. Fast detection of the pupil centre in stable light conditions. In *Proceedings of Central European Seminar on Computer Graphics 2015*, Vienna, Austria, 2015. VUT.
- [YHR67] Alfred L Yarbus, Basil Haigh, and Lorrin A Riggs. *Eye movements and vision*, volume 2. Plenum press New York, 1967.
- [YKLC02] Dong Hyun Yoo, Jae Heon Kim, Bang Rae Lee, and Myoung Jin Chung. Non-contact eye gaze tracking system by mapping of corneal reflections. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 94–99. IEEE, 2002.
- [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 Konceptcje*, volume 3. Wydawnictwo Naukowe PWN, 6 edition, 2009.
- [ZMI99] Shumin Zhai, Carlos Morimoto, and Steven Ihde. Manual and gaze input cascaded (magic) pointing. In *In Proceedings of CHI’99*, pages 246–253, 1999.
- [Zon07] R. Zone. *Stereoscopic Cinema and the Origins of 3-D Film, 1838-1952*. University Press of Kentucky, 2007.
- [ZY02] Jie Zhu and Lei Yang. Subpixel eye gaze tracking. In *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 124–129. IEEE, 2002.