

## RIG Z1.ALG

# Algorytmy analizy i przetwarzania obrazu stereoskopowego

© 2015 FINN Sp. z o.o. Wszelkie prawa zastrzeżone

### Historia zmian dokumentu:

Data	Wersja	Osoba	Opis
2014.06.24	1.0	Przemysław Sztoch Adam Wojciechowski	Utworzenie struktury dokumentu.
2014.09.26	1.1	Jakub Wawrzyniak Agnieszka Rytel	Aktualizacja dokumentu związana z przeprowadzoną weryfikacją całości dokumentacji oraz analizą wymagań.
2014.12.03	1.2	Przemysław Sztoch Ryszard Sztoch Jakub Wawrzyniak	Aktualizacja i weryfikacja wynikająca z przeprowadzonego przeglądu jakości
2015.03.30	1.3	Ryszard Sztoch	Weryfikacja formalna dokumentu – nagłówki i stopki
2015.04.20	1.4	Adam Wojciechowski	Aktualizacja i weryfikacja wynikająca z przeprowadzonych badań i prac rozwojowych
2015.04.29	1.5	Adam Wojciechowski	Aktualizacja i weryfikacja wynikająca z przeprowadzonych badań i prac rozwojowych

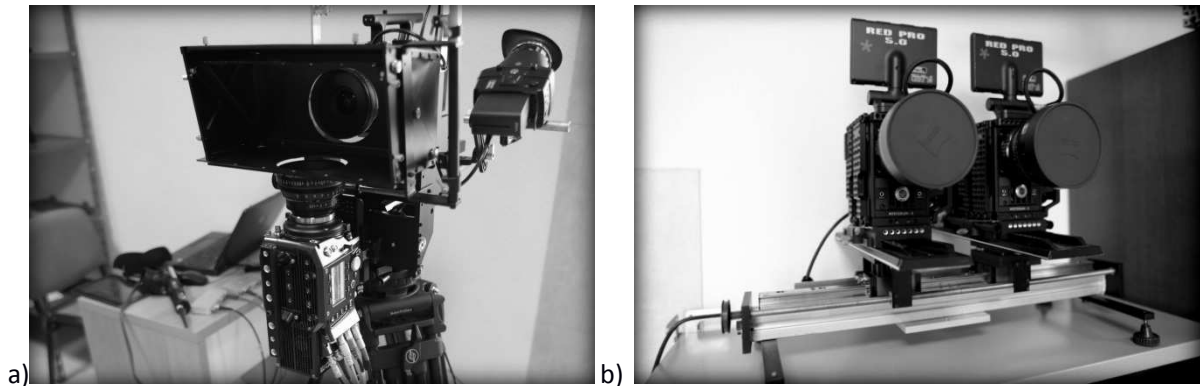
Niniejszy dokument sporządzono na podstawie dokumentacji opracowanej w trakcie realizacji projektu pt. „Rejestracja obrazu stereoskopowego”, realizowanego w ramach Przedsięwzięcia pilotażowego Wsparcie badań naukowych i prac rozwojowych w skali demonstracyjnej „DEMONSTRATOR+”.

Przeprowadzone badania przemysłowe w zadaniu nr 1 wykazały, że podwyższony klatkarz oraz inne bardziej wymagające warunki rejestracji obrazu nie zmniejszają skuteczności działania produktu *Z1.ALG Algorytmy analizy i przetwarzania obrazu stereoskopowego [F, K1]*, który w podstawowym zakresie oparty został na produkcie *Z1.ALG Algorytmy [ROS]*. Dowodem tego jest fakt, że produkt *ROS.KR Kontroler riga [ROS]*, którego właściwe działanie w dużej mierze zależy od jakości omawianych algorytmów działań poprawnie podczas dotychczas wykonanych testów. Wyniki badań pozwalają prognozować, że rozwiązania elektroniczne kontrolera opracowane w projekcie „Rejestracja obrazu stereoskopowego” również sprawdzą się przy współpracy z rgiem ultralekkim.

1	Wewnętrzne i zewnętrzne parametry rigu stereoskopowego.....	4
2	Algorytmy określające zależności pomiędzy parametrami modelu ROS .....	10
3	Algorytmy kalibracji układu kamer stereoskopowych .....	13
3.1	Kalibracja torów optycznych.....	13
3.1.1	Dysparycja pionowa .....	14
3.1.2	Dysparycja pozioma (paralaksa pozioma) .....	14
3.1.3	Rozbieżność pomiędzy akomodacją wzroku a konwergencją obrazów .....	14
3.1.4	Rozbieżności pomiędzy konwergencją a głębią ostrości kamer .....	15
3.1.5	Błędy w jasności/kontraście .....	15
3.1.6	Błędy w ostrości.....	15
3.1.7	Błędy w budźecie paralaksy.....	15
3.1.8	Korekta trapezowości obrazu (ang. keystone effect).....	15
3.1.9	Szereg bezwzględnych miar jakości obrazu.....	15
3.1.10	Prędkość zmian obrazu pomiędzy klatkami .....	15
3.1.11	Obecność obiektów wystających z kadru (szczególnie pierwszoplanowych) .....	15
3.2	Kalibracja chromatyczna.....	16
3.2.1	Błędy w zakresie tonalnym obrazu.....	16
4	Algorytmy budowania mapy głębi (dysparycji) .....	16
4.1	Aktywne metody budowania mapy dysparycji.....	20
4.2	Pasywne metody budowania mapy dysparycji.....	20
5	Bibliografia .....	24

## Wstęp

Rejestracja obrazu stereoskopowego realizowana jest poprzez równoczesną rejestrację dwóch obrazów pochodzących z dwóch kamer umocowanych na rigu. Rig może pozwalać na umocowanie kamer prostopadłe (rys. 1a) lub równoległe (rys. 1.1b).



Rys. 1.1. a) rig prostopadły; b) rig równoległy

Główną różnicą pomiędzy dwoma rigami jest to, że rig prostopadły stosowany jest głównie w sytuacjach, gdy niezbędne jest ustawienie niewielkiej odległości pomiędzy osiami patrzenia kamer (tzw. baza (B)), zaś rig równoległy, gdy minimalna odległość pomiędzy osiami patrzenia kamer jest większa niż szerokość jednej kamery. Tak też rig prostopadły stosuje się głównie dla rejestracji scen z niewielkiej odległości (np. plan filmowy w studio nagrań), zaś rig równoległy dla rejestracji scen planów odległych (np. plenery). Rejestracja obrazu dla riga równoległego jest oczywista (dwie kamery umocowane obok siebie), zaś w przypadku riga prostopadłego obraz rejestrowany jest dla kamery patrzącej na wprost poprzez półprzepuszczalne lustro, zaś w przypadku kamery patrzącej do góry (bądź do dołu) rejestrowany jest obraz odbijający się w lustrze. Półprzepuszczalne lustro powinno być ustawione pod kątem 45 stopni do osi patrzenia obydwu kamer.

Kolejnym parametrem, który odpowiada naturalnej sytuacji postrzegania otoczenia przez człowieka jest konwergencja, czyli zbieżność kierunków patrzenia obojga oczu. W naturalnej sytuacji człowiek, obserwując otoczenie, koncentruje się, co pewien czas na różnych jego elementach. Wtedy kierunki patrzenia obu oczu przecinają się właśnie w punkcie, w którym obserwator „zawiesił wzrok” (ang. fixation point). Ponieważ nasze rzeczywiste otoczenie jest dynamiczne zazwyczaj każdy człowiek koncentruje się, co kilka sekund na innym jego elemencie. W skrajnych sytuacjach możemy zmieniać punkt koncentracji wzroku nawet kilka razy na sekundę. Wraz ze zmianą punktu koncentracji wzroku gałka oczna za pomocą modyfikacji napięcia mięśniowego zmienia swoją geometrię (zmiana ogniskowej -  $f$ ) i wielkość otworu źrenicy (zmiana przysłony -  $A$ ), co skutkuje zmianą głębi ostrości (DoF). Dla obiektów znajdujących się blisko obserwatora głębia ostrości jest mniejsza niż w przypadku obiektów odległych od obserwatora. Ten fizjologiczno-anatomiczny proces wykonywany jest przez człowieka podświadomie, do tego stopnia, że niektórzy nawet nie uświadamiają sobie mechaniki tego procesu. Co więcej, różni obserwatorzy tej samej sceny mogą jednocześnie koncentrować się na innych rzeczach i znajdować się w różnej odległości względem obserwowanych obiektów. Przedstawiony stan rzeczy stawia przed operatorami i reżyserami filmów 3D wyzwanie zaproponowania w czasie rejestracji obrazu konwergencji kamer i głębi ostrości obrazu, która odpowiadałaby „wszystkim” widzom oglądającym film. Jest to zadanie na tyle trudne i wkraczające w sferę psychologii widzenia, że proces dodawania konwergencji jest bardzo często dokonywany na etapie post-produkcji. Zdjęcia na planie filmowym wykonywane są dla równoległe ustawionych kamer (brak konwergencji) przy założonym jedynie (domniemanym) punkcie koncentracji kamer. Parametr ten można zmienić bowiem w post-produkcji.

Parametrami, które ustawia się często na planie filmowym są ogniskowa ( $f$ ) i przysłona ( $A$ ). Na uwagę zasługuje fakt, że najczęściej profesjonalne obiektywy mają stałą ogniskową, przez co zmiana ogniskowej wiąże się ze zmianą obiektywu. Przysłona zaś jest dużo łatwiej modyfikowalnym parametrem i można ją dużo prościej modyfikować wpływając na głębię ostrości.

Osobną kwestią w pracy z rigiem stereoskopowym jest jego kalibracja. Problemem jest, bowiem precyzyjne ustawienie kamer tak, aby ich parametry miały dokładnie takie wartości, jakie wynikają z obliczeń/zamierzeń. Niestety brak oznaczeń na kamerach (brak oznaczonej głównej osi patrzenia kamery) oraz brak precyzji mechanicznego układu nastawnego (luzy na skoku śruby, brak podziałek na większości potencjometrów i suwaków) powoduje, że teoretyczne ustawienia kamery nie odpowiadają praktycznym.

Proces kalibracji zaczyna się zazwyczaj od próby ustawienia układu kanonicznego kamer. Polega on na wyeliminowaniu konwergencji pionowej i poziomej oraz w przypadku rigu prostopadłego na spowodowaniu, aby osie patrzenia dwóch kamer pokrywały się (dla rigu równoległego osie te powinny być jedynie idealnie równoległe). Dla rigu prostopadłego korektę ustawień dokonuje się na podstawie tablicy kalibracyjnej, która pozwala zweryfikować ustawienia kamer w płaszczyźnie widoku.

Dopiero po ustawieniu kamer w układzie ortogonalnym (pot. „wyzerowanie kamer”) asystent operatora kamery ustawia zadane przez reżysera/operatora nastawy rigu. Należy podkreślić, że dla rigu prostopadłego możliwe jest takie ustawienie kamer, w którym widzą one dokładnie ten sam obraz.

Dotychczasowa kalibracja nie pozwala natomiast na skalibrowanie osi patrzenia kamer w 3 wymiarach. Pomijany jest także usuwanie aberracji geometrycznych i chromatycznych obiektywu, które dopiero korygowane są, jeśli w ogóle, na etapie post-produkcji.

Celem dalszej części dokumentu jest wprowadzenie wspólnych oznaczeń i pojęć dla dalszych rozważań, jak również określenie zależności, jakie między tymi parametrami występują. W kolejnych rozdziałach przewidziane jest również wprowadzenie kryteriów poprawności obrazu stereoskopowego oraz metod ich mierzenia

## 1 Wewnętrzne i zewnętrzne parametry rigu stereoskopowego

Wśród zbioru parametrów, które można ustawić dla układu kamer zamocowanych w rigu stereoskopowym można wyróżnić wiele parametrów, jednak najważniejszymi zdają się być grupy parametrów związane z samą kamerą, z rigiem oraz grupa parametrów opisująca zarejestrowany materiał filmowy. W tabeli 1 zebrane są te parametry wraz z ich oznaczeniami i krótkim opisem. Parametry te niezbędne są do jednoznacznego zdefiniowania algorytmów i metod określających relacje pomiędzy parametrami.

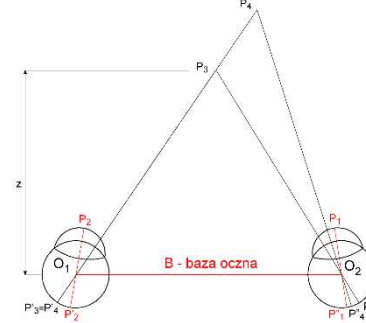
Tab. 1 Zbiór parametrów układu dwóch kamer zamocowanych na rigu stereoskopowym

Nazwa parametru	Symbol	Opis	Uwagi	Przykładowe wartości dla konkretnych rozwiązań
<b>Zbiór parametrów związanych z kamerą</b>				
Identyfikator kamery ( <i>camera id</i> )	ID	Identyfikator kamery, z której było kręcone ujęcie	-	-
Typ kamer	<i>Body type</i>	Producent i model	-	Pracujemy na kamerach Red Epic-X
Szerokość matrycy (mm)	$S_w$	Fizyczna szerokość matrycy w mm	-	Red Epic 5K (25,9 mm) Red One 4K (22,1 mm)

Projekt pt. *Osprzęt i oprogramowanie automatyzujące proces filmowania obrazu 3D* nr Umowy z NCBR: *INNOTECH-K3/H13/16/227587/NCBR/14* dofinansowany w ramach Programu „*INNOTECH*” w ścieżce programowej *HI-TECH*, współfinansowany ze środków Narodowego Centrum Badań i Rozwoju oraz Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego.

( <i>sensor physical width</i> )				Red One 3K (16,75 mm) Red One 2K (11,05 mm) Phantom HD Gold 720 (16 mm) Phantom HD Gold 1080 (16 mm) Phantom HD Gold 1152 (25,6 mm) Phantom Flex 720 (12,8 mm) Phantom Flex 1080 (19,2 mm) Phantom Flex 1140 (25,6 mm)
Wysokość matrycy (mm) ( <i>sensor physical height</i> )	$S_H$	Fizyczna wysokość matrycy w mm		
Współczynnik przycięcia sensora kamery ( <i>sensor crop</i> )	$S_{CROP}$	Współczynnik przycięcia matrycy kamery oblicza się najczęściej w odniesieniu do fizycznego rozmiaru klatki obrazu 35 mm (jej fizyczny rozmiar to 24 mm x 36 mm). Współczynnik ten stosuje się również do obliczania ekwiwalentu ogniskowej.	W niektórych aparatach cyfrowych zamiast posługiwać się współczynnikiem przycięcia wylicza się odpowiednio przeskalowany ekwiwalent ogniskowej, który jest równy $f_{35mm} \times \text{wsp. } S_{CROP}$	Red Scarlet Mysterium-X (36/27.7 = 1,299 CROP) Red Scarlet Dragon (36/25.6 = 1,406 CROP) Red Epic Mysterium-X (36/27.7 = 1,299 CROP) Red Epic 5K (36/25,9=1,3899 CROP)
Współczynnik przycięcia obrazu ( <i>image crop</i> )	$I_{CROP}$	Współczynnik przycięcia kadru biorący pod uwagę rozmiar matrycy w pikselach i rozmiar kadru w pikselach, który wynika z ustawionego formatu zapisu obrazu.		
Rozmiar matrycy w pikselach (pix.) ( <i>image size in pixels</i> )	$S_w \times S_h$	Wymiar matrycy w pikselach		Red Epic (5120(w)x2700(h)) ASPECT (1.896:1)
Współczynnik proporcji obrazu ( <i>aspect</i> )	ASPECT	Proporcje wysokości do szerokości w ramach dostępnych rozmiarów matrycy		2:1 (5120w x 2560h) 2.4:1 (5120w x 2133h) HD16:9 (4800w x 2700h)
<b>Zbiór parametrów związanych z obiektywem</b>				
Ogniskowa obiektywu ( <i>focal length</i> )	$f$	Odległość w milimetrach soczewki od matrycy	Ogniskowa na obiektywie jest modyfikowana przez współczynnik CROP aby obliczyć rzeczywistą wartość ogniskowej, którą realizowane jest ujęcie.	Najczęściej posługujemy się obiektywami stało-ogniskowymi: 18 mm 25 mm 35 mm 50 mm 85 mm 300 mm
Przesłona ( <i>lens aperture</i> )	$A$	Współczynnik określający ograniczenie dopływu światła do matrycy występujący najczęściej	W klasycznej fotografii kolejne wartości przesłony zmniejszały dopływ	f/1.4, f/2, f/2.8, f/4, f/5.6, f/8, f/11, f/16, f/22, f/32, f/45, f/64

		jako dzielnik ogniskowej	światła dwukrotnie. Obecnie progów przesłony jest więcej	
Krążek rozproszenia (mm) - (ang. <i>circle of confusion</i> )	Circle of Confusion <b>CoC</b>	Wielkość plamki rozmycia na matrycy do obliczania głębi ostrości	Przyjmuje się 0,001 cala lub 0,03 mm dla pełnej klatki obrazu 35 mm, dla matryc cyfrowych o innym (najczęściej mniejszym) fizycznym rozmiarze trzeba uwzględnić zmniejszenie tego parametru. Odpowiada on zdolności ludzkiego oka do rozpoznawania szczegółów na poziomie 20 sekund kątowych (przy złym oświetleniu 50-90 sekund kątowych)	
Głębina ostrości (m) (ang. <i>depth of field</i> )	<b>DoF</b>	Głębina ostrości to zakres odległości, w którym obiekty, na zdjęciu sprawiają wrażenie ostrych (rozmazanie punktu nie przekroczy zadanej wartości K (0,001 mm)), czyli tzw. progu ostrego widzenia obrazu. Jeśli K jest większe od zadanego progu obraz uznawany jest jako nieostry	Głębina ostrości zależy od odległości, na której koncentruje się obiektyw (X), ogniskowej obiektywu (f), przesłony (D), wielkości matrycy (WxH)	Dla małej przysłony (duży otwór) rozkład głębi ostrości jest niesymetryczny i mniejszy, lecz dla większych przesłon zakres zaczyna się równoważyć i ogólnie rosnąć. Wraz ze wzrostem odległości fotografowanego obiektu głębina ostrości rośnie to wraz z wydłużeniem ogniskowej maleje (wielkości przeciwstawne).
Poziomy kąt widzenia (°) (ang. <i>horizontal view angle</i> )	$\alpha_{hor}$	Zakres poziomego kąta widzenia wyrażony w stopniach	Czym większa ogniskowa tym mniejszy kąt widzenia	
<b>Zbiór parametrów związanych z rigiem</b>				
Baza (mm) ( <i>stereoscopic base</i> )	<b>B</b>	Odległość pomiędzy osiami patrzenia kamer mierzona w milimetrach w punkcie środka matrycy.		
Konwergencja (°) (ang. <i>convergence</i> )	<b>C</b>	Kąt zbieżności kamer – kąt mierzony w stopniach, jaki tworzą dwie osie kamer w punkcie przecięcia ich osi patrzenia	Zazwyczaj film kręcony jest bez konwergencji a ona dokładana jest na etapie postprodukcji	

<p>Paralaksa (%) (ang. <i>parallax</i>)</p>	<p><b>P(z)</b> <b>P(z<sub>near</sub>)</b> dla punktów odległych o <b>z<sub>near</sub></b> od kamery <b>P(z<sub>far</sub>)</b> dla punktów odległych o <b>z<sub>far</sub></b> od kamery <b>P(z<sub>screen</sub>)=0</b></p>	<p>Odległość kątowna lub liniowa pomiędzy odpowiadającymi sobie punktami na dwóch stereo receptorach wyrażona odpowiednio w stopniach lub w milimetrach. Rozbieżność w procentach pomiędzy wybranymi elementami kadru w nałożonych na siebie obrazach lewego i prawego oka. Najistotniejsze z punktu widzenia obrazu stereoskopowego jest <b>paralaksa dodania (P(z<sub>far</sub>))</b> dla obrazów widocznych w najdalszym planie kadru oraz <b>paralaksa ujemna (P(z<sub>near</sub>))</b> dla obiektów pierwszoplanowych. Paralaksa jest zerowa w płaszczyźnie, gdzie obrazy dla obu oczu się pokrywają (<b>płaszczyzna zerowej paralaksy P(z<sub>screen</sub>)=0</b>;</p>	<p>Wartość paralaksy mierzona w milimetrach dla docelowego ekranu projekcyjnego nie powinna przekraczać odległości oczu widza (standardowo przyjmuje się 65 mm). Ponieważ możemy nie znać dokładnie fizycznej szerokości ekranu projekcyjnego to paralaksę podaje się w procentach, a wtedy paralaksę dla danego ekranu projekcyjnego można policzyć:</p> <p><b>SZER_EKR(cm) * P<sub>budget</sub>(%) = P<sub>budget</sub> (cm)</b></p> <p>Operatorzy na planie operują sumaryczną paralaksą <b>P(z<sub>near</sub>) + P(z<sub>far</sub>) = P<sub>budget</sub> – Budżet Paralaksy</b> (taką paralaksę widać na ekranie podglądu na planie filmowym), a ich wzajemny udział/podział uzyskuje się w postprodukcji, przesuwając obrazy z dwóch kamer względem siebie.</p>	<p>Dla ekranów kinowych przyjmuje się <b>P(z<sub>near</sub>)=2%</b> i <b>P(z<sub>far</sub>)=0.7%</b> <b>0.7% (P(z<sub>far</sub>) w %) * 1000 cm (SZER_EKR 10 m) = 7 cm (P(z<sub>far</sub>) w cm)</b></p>
	<p>Paralaksa fizjologiczna</p>	 <p><math>O_1O_2 = B</math> baza oczna  <math>Z</math> odległość punktu <math>P_3</math> od bazy ocznej  <math>\angle O_1P_3O_2</math> kąt paralaktyczny pod jakim obserwowany jest punkt <math>P_3</math>  <math>\angle O_1P_3O_2 - \angle O_1P_4O_2</math> miara zdolności rozdzielczej widzenia stereoskopowego  <math>P_2P_3 - P_1P_4</math> paralaksa fizjologiczna - różnica długości łuków    usytuowanie garek ocznych przy obserwacji punktu.    Miara zdolności rozdzielczej widzenia stereoskopowego</p>		
<p>Odległość od obiekt-u/-ów pierwszoplanowego (m) (ang. <i>near plane distance</i>)</p>	<p><b>Z<sub>near</sub></b></p>	<p>Odległość najczęściej podawana w metrach</p>	<p>Parametr niezbędny do ustawień kamer i kalkulatorów najczęściej rejestrowany za pomocą laserowego dalmierza przez asystenta kamerzysty. Parametr</p>	<p>Istnieje reguła, że pierwszy plan (obiekt) nie powinien być bliżej niż 30 krotność Bazy (<b>B</b>), lub nawet niektórzy podają, że 50 krotność Bazy.</p>

			niezbędny (najczęściej służebny, gdyż łatwo go zmienić) do weryfikacji ustawień kamer (np.: ogniskowa, głębia ostrości, itp.).	
Odległość od obiekt-u/-ów z dalekiego planu (m) (ang. <i>far plane distance</i> )	<b>Z_far</b>	Odległość najczęściej podawana w metrach	Parametr niezbędny do ustawień kamer i kalkulatorów najczęściej rejestrowany za pomocą laserowego dalmierza przez asystenta kamerzysty, lub znany na podstawie wielkości planu filmowego.	
Odległość od płaszczyzny zerowej paralaksy (m) (ang. <i>screen plane distance</i> )	<b>Z_screen</b> $P(z_{screen})=0$	Odległość od kamery, w której obrazy lewego oka i prawego oka pokrywają się.	W przypadku układu kamer zbieżnych wielkość ta wynika z poziomej konwergencji kamer. Dla rejestracji obrazu z układu kamer równoległych wielkość ta jest jedynie wstępnie zakładana gdyż przy rejestracji oba obrazy pokrywają się jedynie w nieskończoności. Na etapie post-produkcji uzyskuje się pokrycie obrazów lewej i prawej kamery (efekt sztucznie stworzonej konwergencji) przesuając klatki obrazu lewej i prawej kamery względem siebie (w poziomie). Przesunięcie pionowe jest niedopuszczalne.	Planowana konwergencja przy równoległej rejestracji obrazów wynika pośrednio z budżetu paralaksy, który powinien być podzielony w takich proporcjach na paralaksę ujemną i dodatnią, aby nie przekroczyć niedozwolonych poziomów.
Punkt koncentracji ostrości obiektywu (m) (ang. <i>focus distance</i> )	<b>Z_focus</b> ( $z_f$ )		Odległość od kamery, której skoncentrowana jest ostrość widzenia obiektywu. Od tego dystansu do przodu i do tyłu zawiera się przestrzeń głębi ostrości widzenia	
Post-produkcyjna konwergencja pozioma (%) (ang. <i>postproduction horizontal convergence</i> )	<b><math>\Delta x = P(z_{near})</math></b>	Przesunięcie obrazów lewej i prawej kamery względem siebie powodujące post-produkcyjną konwergencję, czyli ustawienie w planie ostrym obiektów odległych o <b>z_screen</b> od kamery.	Na etapie post-produkcji obrazy lewej i prawej kamery należy przesunąć o $\Delta x$ pikseli (wyrażone w liczbie pikseli lub procentowo w funkcji szerokości kadru), aby obrazy z dwóch kamer pokryły się w planie odległym o <b>z_screen</b>	Przesunięcie obrazów względem siebie powoduje obcięcie kadru z lewej i prawej strony, które powinno być sygnalizowane operatorowi na podstawie $\Delta x$ .



			(odległość płaszczyzny zerowej paralaksy) od kamer.	
Post-produkcyjna konwergencja pionowa (%) (ang. <i>postproduction horizontal convergence</i> )	<b>Δy</b>	Przesunięcie obrazów lewej i prawej kamery względem siebie w pionie, powodujące post-produkcyjną konwergencję pionową, jest efektem niepożądanym.	Pionowa konwergencja jest efektem niepożądanym i jeśli występuje to jest mierzona w promilach (%). Może pojawić się zarówno podczas rejestracji przy pomocy rigu, na stole montażowym jak i podczas projekcji, gdy dwa obrazy (rzutniki) nie zostaną ze sobą odpowiednio spasowane. Parametr ten stanowi jedno z kryteriów poprawności obrazu stereoskopowego. Pionowa konwergencja może być źródłem złego samopoczucia widzów oglądających obraz 3D.	
<b>Zbiór parametrów związanych z klipem filmowego</b>				
Identyfikator kamery	<b>camera id</b>			
Identyfikator karty	<b>reel id</b>			
Identyfikator klipu	<b>clip id</b>			
Data nagrania	<b>date</b>			
Czas nagrania	<b>Abs TC</b>			
Rzeczywisty czas nagrania	<b>End Abs TC</b>			
Długość nagrania				
Całkowita liczba klatek				
Rozdzielczość nagrania w pikselach				
Aspekt nagrania				
Bazowy klatkarz	<b>Base framerate</b>			
Klatkarz nagrania				
ISO nagrania				
Tryb HDR				
Współczynnik kompresji obrazu		Współczynnik pozwalający określić jaki długi materiał filmowy zmieści się na dysku.		

Głębina bitowa pikseli obrazu				
Rozkład jasności w scenie				

Parametry te ustawiane są zazwyczaj na dwóch etapach obsługi rigu i niektóre na etapie post-produkcji.

Pierwszy etap dotyczy kalibracji kamer, czyli wyzerowania ustawień układu kamer w taki sposób, aby dwie kamery rejestrowały dokładnie ten sam obraz (ten sam kadr, ta sama kolorystyka, jasność, ostrość, te same aberracje geometryczne, winietowanie i in.). Drugi etap to modyfikacja tych parametrów (np.: baza, ogniskowa, przesłona), które pozwalają uzyskać zamierzony efekt artystyczny (głębokość ostrości, głębię stereoskopową, kadr, itp.). Trzeci etap – post-produkcji pozwala zmodyfikować te parametry, które wskazują na niezamierzone różnice ustawień pomiędzy dwoma kamerami w rigu (np.: różnica jasności, ostrości) oraz te, które wynikają z zamierzeń artystycznych operatora czy reżysera (np.: konwergencja, paralaksa).

Drugi etap ustawień zestawu kamer w rigu stereoskopowym odbywa się głównie na planie zdjęciowym i polega w głównej mierze na modyfikacji następujących parametrów:

- baza ( $B$ )
- ogniskowa obiektywu ( $f$ )
- przesłona ( $A$ )
- ewentualnie konwergencja ( $C$ ), jeśli jest ustawiana
- pozycja kamery względem obiektów planu ( $z_{near}$ ,  $z_{far}$ )

Etap post-produkcyjny polega zazwyczaj na wykonaniu korekt barwnych obrazu, korektach jasności oraz przede wszystkim na wprowadzeniu post-produkcyjnie konwergencji poprzez wzajemne przesunięcie w poziomie obrazów względem siebie ( $\Delta x$ ). Przesunięcie to z definicji powinno być równe założonej paralaksie ujemnej wyrażonej w procentach ( $P(z_{near})$ ), jednak decyzją operatora może zostać zmienione dla uzyskania lepszych efektów stereoskopowych. Post-produkcyjna konwergencja pionowa ( $\Delta y$  wzajemne przesunięcie obrazów w pionie) jest mierzona w promilach i jest efektem niepożądanym lub wręcz jest błędem pogarszającym odbiór obrazu stereoskopowego.

## 2 Algorytmy określające zależności pomiędzy parametrami modelu ROS

Jedną z pierwszych zasad podczas rejestrowania obrazu 3D jest reguła określania odległości kamery od pierwszego planu. Zasada uzależniająca odległość pierwszego planu ( $z_{near}$ ) w zależności od przyjętej Bazy rigu ( $B$ ) dana jest wzorem 3.1. Wynika z niego, że odległość do pierwszego planu ( $z_{near}$ ) powinna być, co najmniej 30-krotnością Bazy ( $B$ ).

$$B = \frac{z_{near}}{30} \quad (3.1)$$

gdzie  $B$ - baza kamer,  $z_{near}$ - odległość od najbliższego obiektu.

Niektórzy podają, że dla bardziej odległych obiektów (rzędu 10 metrów) bazę należy liczyć według wzoru 3.2, gdzie minimalna odległość jest 50-krotnością Bazy ( $B$ ), a niektórzy uznają za dobrą praktykę ustawianie pierwszego planu na dystansie 100-krotności Bazy ( $B$ ).

$$B = \frac{z_{near}}{50} \quad (3.2)$$

Jest to zbieżne z teorią, która mówi, że idealny odbiór efektu stereoskopowego następuje przy zbieżności osi patrzenia rzędu 1<sup>o</sup>-2<sup>o</sup>.

Kolejną operacją wykonywaną dość często podczas realizacji obrazu stereoskopowego jest obliczanie Bazy (**B**) dla przyjętej wielkości budżetu Paralaksy ( $P_{budget}$ ) wyrażonej w procentach (%) i dla układu równoległych kamer z nieskończonością na obrazie. Wielkość paralaksy w mm oblicza się na podstawie przyjętego budżetu paralaksy w procentach i wielkości (szerokości) ekranu, na którym wyświetlany będzie docelowo obraz. Wzór 3.3 pozwala na przeliczenie wielkości procentowych paralaksy  $P$  na wielkości wyrażone w mm.

$$P_{budget(mm)}(z) = W_{screen} \cdot P_{budget(\%)}(z) \quad (3.3)$$

gdzie  $P_{budget(mm)}$  paralaksa w mm,  $W_{screen}$  – fizyczna szerokość ekranu w mm,  $P_{budget(\%)}$  - paralaksa wyrażona w %. Wartość paralaksy ujemnej nie powinna przekroczyć 65 mm, ile wynosi średni odstęp pomiędzy oczami widza. Szczególnie paralaksa dodatnia dla planu dalekiego nie powinna powodować, że oczy obserwatora będą rozbieżne.

Uproszczony wzór obliczenia bazy (**B**) dla przyjętej paralaksy ( $P$ ), ogniskowej ( $f$ ) i odległości od planu  $z$ , dla którego paralaksę liczymy, wyrażony jest wzorem 3.4.

$$\frac{B}{z} = \frac{P}{f} \quad \text{gdzie} \quad z_{near} \leq z \leq z_{far} \quad (3.4)$$

W uproszczonym podejściu przyjmuje się dla pełnej klatki 35 mm wielkość paralaksy wynoszącą 1.2 mm i wtedy wzór 3.4 przyjmuje uproszczoną postać 3.5 dla naturalnych jednostek ( $z$  w metrach,  $f$  w milimetrach)

$$B = \frac{1200 \cdot z}{f} \quad \text{gdzie} \quad z_{near} \leq z \leq z_{far} \quad (3.5)$$

Dużo wygodniejsze i bardziej zbieżne ze spotykanymi w rzeczywistości kalkulatorami jest posługiwanie się wzorem 3.6, który będzie łączył Bazę i Paralaksę dla układu kamer równoległych z głębią ostrości w zakresie od  $z_{near}$  (najbliższy obiekt) do  $z_{far}$  (najdalszy obiekt) dla *camera obscura*.

$$\begin{aligned} B &= \frac{[P(z_{near}) + P(z_{far})] \cdot W}{f \left( \frac{1}{z_{near}} - \frac{1}{z_{far}} \right)} = \frac{[P(z_{near}) + P(z_{far})] \cdot W \cdot z_{far} \cdot z_{near}}{f (z_{far} - z_{near})} = \\ &= [P(z_{near}) + P(z_{far})] \cdot W \cdot \frac{z_{far} \cdot z_{near}}{(z_{far} - z_{near})} \cdot \frac{1}{f} \end{aligned} \quad (3.6)$$

gdzie **B**- baza układu kamer, **f** - ogniskowa obiektywu,  $z_{near}$  – odległość od najbliższego obiektu planu,  $z_{far}$  – odległość od najdalszego obiektu planu,  $P(z)$  – paralaksa dana wzorem 3.3, przy czym Paralaksa wyrażona w procentach jest budżetem paralaksy ( $P_{budget(\%)}$ ), czyli sumą paralaksy ujemnej i dodatniej.

Dodatkowo chcąc ustawić na planie konwergencję kamer możemy ją obliczyć na podstawie procentowego udziału Paralaksy dodatniej i ujemnej w zakresie przyjętego budżetu paralaksy. Odległość ( $z_{screen}$ ) do punktu zbieżności kamer można policzyć na podstawie wzoru 3.7

$$z_{screen} = \frac{f \cdot B}{\frac{f \cdot B}{z_{near}} - P(z_{near}) \cdot W} \quad (3.7)$$

zaś konwergencję kamer (**C**) wyrażoną w stopniach można policzyć ze wzoru 3.8.

$$C = 2 \cdot \arctg\left(\frac{\frac{B}{2}}{z_{screen}}\right) \quad (3.8)$$

W literaturze można spotkać modyfikację wzoru 3.6. Dla kamer z obiektywem wzór 3.6 może przybrać kształt reguły Berkovitz'a (wzór 3.9).

$$\begin{aligned} B &= P \cdot \frac{z_{far} z_{near}}{(z_{far} - z_{near})} \cdot \left( \frac{1}{f} - \frac{z_{far} + z_{near}}{2z_{far} z_{near}} \right) = P \cdot \frac{z_{far} z_{near}}{(z_{far} - z_{near})} \cdot \left( \frac{2z_{far} z_{near}}{2fz_{far} z_{near}} - \frac{f(z_{far} + z_{near})}{2fz_{far} z_{near}} \right) = \\ &= P \cdot \frac{(2z_{far} z_{near} - f(z_{far} + z_{near}))}{2f(z_{far} - z_{near})} \end{aligned} \quad (3.9)$$

gdzie

$$P = [P(z_{near}) + P(z_{far})] \cdot W$$

Dla ogniskowej hiperfokalnej (głębia ostrości w nieskończoności) reguła Berkovitz'a przyjmuje uproszczoną postać (wzór 3.10)

$$B = \frac{P \cdot z_{near}}{\left(f - \frac{1}{2}\right)} \quad (3.10)$$

Davis zaproponował modyfikację formuły Berkovitz'a jeśli  $z_{far} < 2 \cdot z_{near}$  to wtedy przyjął, że  $z_{far} = 2 \cdot z_{near}$ . Jednakże większość kalkulatorów bazuje na wzorze 3.6.

Kolejnym parametrem, który może interesować operatora jest kąt widzenia obiektywu dla przyjętej ogniskowej (**f**) i wielkości matrycy (**S<sub>w</sub>xS<sub>H</sub>**). Wartość poziomego kąta widzenia obiektywu (**α<sub>hor</sub>**) przy ustalonej fizycznej szerokości matrycy (**S<sub>w</sub>**) dana jest wzorem 3.11.

$$\alpha_{hor} = 2 \cdot \arctg\left(\frac{S_w}{f}\right) \quad (3.11)$$

gdzie **S<sub>w</sub>** to fizyczna szerokość matrycy, **f** – ogniskowa obiektywu.

Innym aspektem, który jest analizowany na planie filmowym jest głębia ostrości poszczególnych kamer. Empirycznie obraz uznawany jest za ostry (wyraźny), jeśli tzw. krążek rozmycia obrazu (**CoC**) nie przekracza 0.03 mm lub 0,001 cala dla pełnej klatki obrazu (35 mm). Jeśli matryca rejestracji obrazu jest fizycznie mniejsza od pełnej ramki obrazu to odpowiednio należy też zmniejszyć krążek rozmycia obrazu, gdyż ważne jest przede wszystkim ostrość obrazu na wyświetlaczu (ekran, monitor).

Dla przyjętego punktu koncentracji ostrości (ang. *focus distance*) (**z<sub>focus</sub>**), ogniskowej obiektywu (**f**) i przysłony (**A**) ogólny wzór określający maksymalną i minimalną odległość, przy której obraz jest wyraźny określa wzór 3.12 i 3.13.

$$z_{focus\ near} = \frac{f^2 \cdot z_{focus}}{f^2 - CoC \cdot A \cdot (a \cdot z_{focus} - f)} \quad (3.12)$$

$$z_{focus\ far} = \frac{f^2 \cdot z_{focus}}{f^2 + CoC \cdot A \cdot (a \cdot z_{focus} - f)} \quad (3.13)$$

gdzie  $a$  jest współczynnikiem proporcji wielkości matrycy w odniesieniu do pełnej ramki obrazu (zazwyczaj  $<1$ ). Głębina ostrości jest wtedy dana wzorem 3.14.

$$DoF = z_{focus\ far} - z_{focus\ near} \quad (3.14)$$

## 3 Algorytmy kalibracji układu kamer stereoskopowych

### 3.1 Kalibracja torów optycznych

Tworzenie poprawnego obrazu stereoskopowego jest zagadnieniem, które wymaga uwzględnienia szeregu czynności zarówno na etapie przygotowania (preprodukcja), jak również na etapie realizacji (produkcja) i postprodukcji. Problemy w jakości obrazu mogą leżeć zarówno w dopasowaniu parametrów kamer jak i w samej jakości obrazu (kadr, szybkość prowadzenia kamer). Dopasowanie parametrów kamer realizowane jest zarówno na etapie wstępnym (kalibracja) jak i podczas realizacji filmowej. O ile na etapie przygotowania obsługa ma stosunkowo dużo czasu na skalibrowanie układu kamer, to na etapie realizacji tego czasu jest relatywnie mało.

Istotnym elementem na etapie przygotowania jest kalibracja ustawień riga.



Rys. 4.1. a) rig prostopadły; b) rig równoległy

Rozbieżności niektórych parametrów dla obrazów lewego i prawego oka jest bardzo trudno wyeliminować lub jest to niemożliwe.

Dotyczy to w szczególności riga lustrzanego, w którym kamery zazwyczaj są ustawione prostopadle. Wynika to z tego, że jeden obraz (zazwyczaj oka lewego) jest rejestrowany jako światło przepuszczone przez półprzepuszczalne lustro (zazwyczaj przez kamerę poziomą). Drugi obraz (oko prawe) jest rejestrowany jako światło odbite od półprzepuszczalnego lustra (zazwyczaj przez kamerę pionową). Wpływa to na różnice jasności, różnice w zakresie tonalnym oraz różne zabarwienie obu obrazów.

W przypadku rigu równoległego różnice mogą wynikać z różnego ustawienia kamer względem obecnych w scenie źródeł światła. Ponieważ rig równoległy jest zazwyczaj używany na otwartych przestrzeniach słońce jest bezpośrednią przyczyną powstawania refleksów i rozbieżności.

W trakcie wykonywania zdjęć na planie filmowym cały czas istnieje potrzeba korygowania nastaw kamer i riga, aby odpowiadały one aktualnie nagrywanym ujęciom. Wykonywane na bieżąco zmiany nastaw (np. baza stereoskopowa, konwergencja, ogniskowa, przysłona, ostrość, klatkarz nagrywania, czas ekspozycji)

Projekt pt. *Osprzęt i oprogramowanie automatyzujące proces filmowania obrazu 3D* nr Umowy z NCBR: *INNOTECH-K3/HI3/16/227587/NCBR/14* dofinansowany w ramach Programu „*INNOTECH*” w ścieżce programowej *HI-TECH*, współfinansowany ze środków Narodowego Centrum Badań i Rozwoju oraz Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego.

powinny ze sobą współgrać i spełniać odpowiednią korelację. Nieodpowiednie ustawienie bazy stereoskopowej względem planu zdjęciowego może skutkować niewłaściwą dysparcją obrazów (paralaksą) podczas projekcji, co będzie powodowało znaczące obniżenie komfortu odbioru obrazu stereoskopowego. Podobnie częste przestawianie parametrów riga oraz jego przenoszenie może powodować utratę kalibracji wertykalnej skutkującej dysparcją pionową, która również nie jest pożądana.

Niewłaściwe rozbieżności mogą pojawić się na polu dopasowania głębi ostrości obiektywów, która powinna zawierać się w głębi obrazu stereoskopowego.

Ewentualne błędy mogą pojawić się w obszarze rozbieżności akcji na planie filmowym, a obszarem koncentracji układu kamer. W tym aspekcie mamy, bowiem do czynienia z sytuacją, kiedy to reżyser/operator planuje i realizuje określony scenariusz chcąc pokazać widzowi pewną scenę. Nie powinna pojawić się wtedy rozbieżność pomiędzy punktem kadru, w którym rozgrywa się akcja a punktem kadru, w którym koncentrują się kamery (mają ustawioną: linię optyczną, głębię ostrości, głębię obrazu).

Błędy w percepcji obrazu stereoskopowego mogą pojawić się również, gdy szybkość zmian na ekranie będzie ponadnaturalna, tak więc prędkość zmian położenia/orientacji rigu w rejestracji obrazu stereoskopowego powinna być niższa niż przy rejestracji obrazu 2D.

Cześć z błędów w obrazie stereoskopowym może powstać również na etapie postprodukcji. Należą do nich głównie nieskoordynowane zmiany jasności, kontrastu, zakresu tonalnego, dysparcji poziomej, pionowej itp. Jednakże większość z tych błędów jest odwracalna, w ramach historii projektu, pod warunkiem świadomości ich popełnienia.

### 3.1.1 Dysparcja pionowa

Pionowa dysparcja (przesunięcie) obrazów powinna być całkowicie zlikwidowana. W rzeczywistości odpowiada to sytuacji, w której widz ma jedno oko wyżej, a drugie niżej. Błąd ten może zostać usunięty w postprodukcji i może przykładowo wynikać ze złego skalibrowania kamer, niewłaściwego nałożenia obrazów w postprodukcji lub nawet niestarannego spasowania dwóch rzutników podczas projekcji w kinie. Błąd ten dość prosto daje się skorygować w postprodukcji. Niemniej powinniśmy dołożyć niezbędnej staranności aby wyeliminować ten błąd już w trakcie samej rejestracji (na etapie produkcji) – z pewnością będzie to tańsze i szybsze.

### 3.1.2 Dysparcja pozioma (paralaksa pozioma)

Zbyt duża pozioma dysparcja (przesunięcie większe niż ok. +6,5 cm) powoduje, że oczy widza (kierunek patrzenia każdego oka) oglądającego materiał ustawiają się rozbieżnie. Wpływa to na złe samopoczucie, nudności i zawroty głowy. Aby zapewnić prawidłową zbieżność oczu podczas oglądania należy podczas rejestracji obrazu zapewniać odpowiedni budżet paralaksy obrazu i na etapie postprodukcji odpowiedni podział pomiędzy paralaksę ujemną (bliską) i dodatnią (daleką).

### 3.1.3 Rozbieżność pomiędzy akomodacją wzroku a konwergencją obrazów

Podczas oglądania filmu widz kieruje się emocjami i koncentruje wzrok na elementach, które z jego punktu widzenia (wynikającym z fabuły) są najistotniejsze. Najczęściej jest to zgodne z wizją reżysera/operatora, który koncentruje kadr na najistotniejszych, z punktu widzenia fabuły, elementach. W przypadku obrazu 3D należy dodatkowo zadbać, aby obiekty istotne z punktu widzenia fabuły (jak również przypuszczalnie widza) znajdowały się na płaszczyźnie zerowej paralaksy, gdzie dzięki konwergencji krzyżują się osie patrzenia kamer (w centrum ostrości obrazu). W skrajnym przykładzie przy rozbieżności akomodacji i konwergencji widz koncentruje się na „nieostrym” fragmencie obrazu, zaś ostrość jest ustawiona na mniej istotnych, z punktu widzenia fabuły, elementach.

### 3.1.4 Rozbieżności pomiędzy konwergencją a głębią ostrości kamer

Ustawiając lub planując konwergencję kamer w Rigu 3D niejako zakładamy, w jakim zakresie odległości od kamer obiekty planu będą mieściły się w budżecie paralaksy – będą widziane maksymalnie z założonym stopniem rozmycia. Wartość ta nakłada się na ustawienia samej kamery (ogniskowa, przysłona itp.), które skutkują określoną głębią ostrości każdego obrazu z osobna. Nie dość, że głębia ostrości każdej z kamer powinna być taka sama to jeszcze powinna ona pokrywać się ostrością stereoskopową wynikającą z konwergencji obrazów.

### 3.1.5 Błędy w jasności/kontraście

Obrazy pochodzące z dwóch kamer powinny mieć odpowiednio identyczną (z pewną tolerancją) ogólną jasność oraz kontrast obrazu. Różnica w jasności lub kontraście może wynikać przykładowo z ustawienia źródeł światła na planie lub samej konstrukcji riga (np. jeden obraz jest rejestrowany przez półprzepuszczalne lustro, a drugi jako obraz odbity od tego lustra).

### 3.1.6 Błędy w ostrości

Jednym z błędów ostrości, poza różnymi ustawieniami kamer może być koncentracja ostrości symetryczna bądź asymetryczna, jak również sama ostrość i jej głębia. Ten błąd w niewielkim zakresie daje się skorygować w postprodukcji.

### 3.1.7 Błędy w budżecie paralaksy

Jeśli budżet paralaksy jest zbyt duży to bardzo trudno jest wyeliminować błędy paralaksy za pomocą samej zmiany proporcji pomiędzy paralaksą ujemną a paralaksą dodatnią.

### 3.1.8 Korekta trapezowości obrazu (ang. keystone effect)

Obrazy stereoskopowe rejestrowane przy ustawionej konwergencji kamer uwidaczniają nagrywaną scenę z nieco innej perspektywy. Technicznie skutkuje to zjawiskiem, w którym pierwszoplanowe obiekty umieszczone na skraju kadru mogą na dwóch obrazach mieć nieznacznie różną wielkość. Zjawisko to można eliminować na etapie postprodukcji stosując wyprostowanie obrazów (tzw. rektyfikacja), a następnie kadrując dodatkowo ramki obrazu (część wspólna dwóch kadrów).

### 3.1.9 Szereg bezwzględnych miar jakości obrazu

Istnieje szereg miar jakości obrazu 2D, z których niektóre okazują się jako mające wysoką korelację z subiektywnymi odczuciami widza. Należą do nich VQM (Video Quality Metric) [1], SSIM, UQI, C4, RRIQA [2], *binocular distortions* [3,4,5,6]

### 3.1.10 Prędkość zmian obrazu pomiędzy klatkami

Parametr ten ma znaczenie z punktu widzenia percepcji widza. Punktem odniesienia jest szybkość obserwowalnych przez człowieka zmian zachodzących w naturze, która pomimo teoretycznie ograniczonych prędkością światła wartości jest znacznie niższa. Istnieje dla każdej osoby pewien subiektywny próg, powyżej którego zmiany zachodzą za szybko i percepcja (rozumienie) zachodzących zmian jest ograniczone. Bazując na opinii ekspertów/badaniach na osobach można pokusić się o wskazanie granicznych wartości progów zmian zachodzących dla poszczególnych planów, powyżej których obraz zaczyna być „drażniący”.

### 3.1.11 Obecność obiektów wystających z kadru (szczególnie pierwszoplanowych)

W sytuacji, gdy pierwszoplanowy obiekt wystaje z kadru (jest częściowo widoczny) obserwator ma problem z pełnym określeniem głębi danego obiektu. Najczęściej głębia obiektu w scenie kłóci się z głębią fizyczną ekranu, na którym oglądany jest film. Poza tym, jeśli obiekt wchodzi w kadr tylko jednej kamery, to znaczy, że będzie widoczny tylko przez jedno oko obserwatora i wtedy będzie mu się on wydawał przezroczysty.

Jeżeli obraz rejestrujemy kamerami ustawionymi równolegle (brak konwergencji) to musimy być świadomi o konieczności skorygowania konwergencji w postprodukcji. Będzie to wiązało się z koniecznością dodatkowego przycięcia obrazów, co zwiększa ryzyko, że obiekty z bliskiego planu będą widziane tylko „jednym okiem”.

## 3.2 Kalibracja chromatyczna

### 3.2.1 Błędy w zakresie tonalnym obrazu

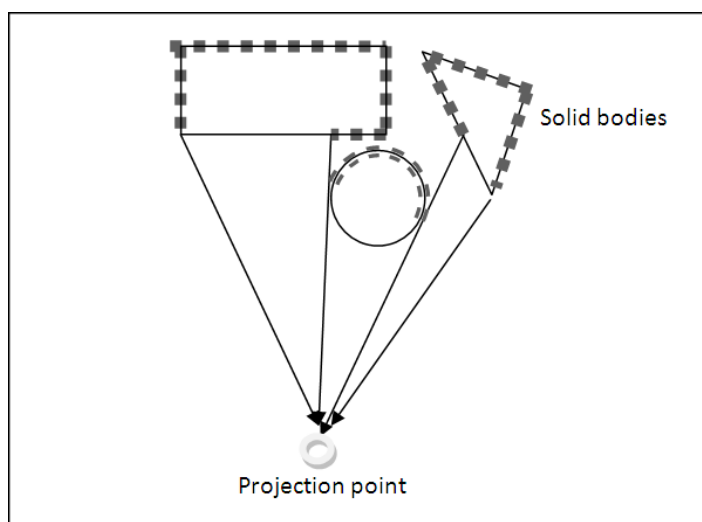
Jeśli z niewyjaśnionych przyczyn zakres tonalny obrazów pochodzących z dwóch kamer się różni to ciężko jest skorygować te błędy, chyba że znamy/znajdziemy korekcję barwną ujednociającą obrazy.

## 4 Algorytmy budowania mapy głębi (dysparycji)

Budowanie mapy dysparycji jest swego rodzaju szacowaniem głębokości poszczególnych elementów obrazu na podstawie najczęściej dwóch obrazów pochodzących z dwóch osobnych kamer ustawionych w znanej względem siebie relacji. Można również rekonstruować głębię na podstawie światła strukturalnego [Sch03].

Obrazy poszczególnych kamer są dwuwymiarową reprezentacją scen trójwymiarowych (projekcją perspektywiczną na płaszczyznę rzutni). Każdy obraz jest z definicji pozbawiony 3 wymiaru (głębokości). Chociaż istnieją algorytmy, które próbują na podstawie 2-wymiarowego obrazu oszacować głębokość poszczególnych punktów obrazu (rekonstrukcja utraconej informacji o głębokości) to dużo dokładniejsze oszacowanie bazujące na dwóch obrazach, na wzór systemu wzrokowego człowieka, który szacuje odległość na bazie obrazów zarejestrowanych przez parę oczu o stałej odległości.

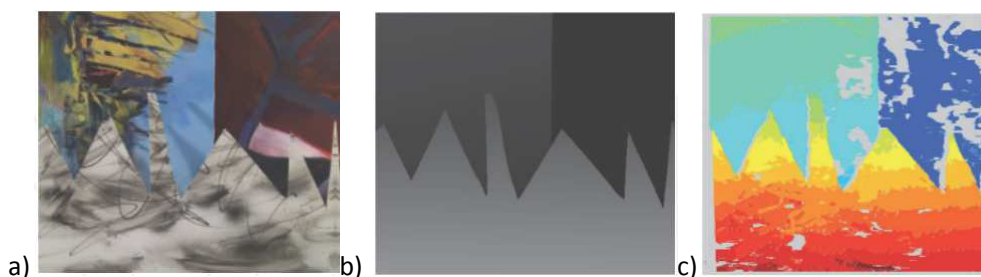
Probleмами przy rekonstrukcji głębi jest m.in. zbiór punktów znajdujących się w nieskończoności (dalekie tło), których głębokość trzeba oszacować oraz w przypadku rekonstrukcji geometrycznej obiektów problem może stanowić zbiór punktów zasłoniętych, których nie widać na obrazie/-ach (rys. 5.1)



Rys. 5.1 zbiór punktów niewidocznych oznaczony jest kwadratami

Kodowanie głębi najczęściej obrazowane jest za pomocą obrazu w skali szarości, gdzie jasność poszczególnych pikseli reprezentuje odległość – czym jaśniejszy piksel tym bliżej kamery (rys. 5.2b). Można również spotkać kodowanie głębi za pomocą koloru [Kos03, Sax08]. Wtedy przykładowo kolor ciemno czerwony reprezentuje obiekty najbliższe, zaś ciemno fioletowy obiekty najdalsze (rys. 5.2c).





Rys. 5.2 Przykłady kodowania głębokości a) rysunek oryginalny, b) skalą szarości, c) kolorem

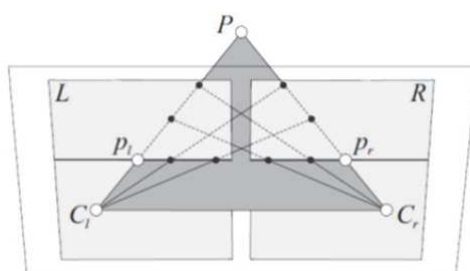
Wspomniane powyżej sposoby kodowania głębi pozwalają nie tylko wizualizować w sposób charakterystyczny mapę głębi, ale również pozwalają porównywać wyniki działania różnych algorytmów. Takie podejście ma miejsce w uznawanej w wizji stereoskopowej bazie danych Middlebury [Sch10]. Baza Middlebury zawiera obrazy stereo pary i odpowiadającą im mapę głębi (rys. 5.3).



Rys. 5.3. lewy obraz (po lewej) i prawy obraz (w środku) oraz odpowiadająca im mapa głębi (po prawej)

O ile wynik demonstrujący mapę głębi może być przedstawiony w skali szarości o tyle same obrazy źródłowe, które służyły do jej obliczania powinny być raczej kolorowe niż w skali szarości, gdyż obraz kolorowy niesie w sobie dużo więcej informacji niż obraz jedynie z poziomami jasności. Z drugiej strony obraz kolorowy jest dużo trudniejszy do przetwarzania ze względu na rozmiar plików i przy algorytmach, w których istotna jest wydajność przetwarzania może to stanowić problem.

Zagadnienie budowania mapy głębi najprościej zrealizować w kanonicznym układzie kamer, w którym kamery patrzą dokładnie równolegle, obrazy z dwóch kamer są współpłaszczyznowe i nie są względem siebie obrócone.

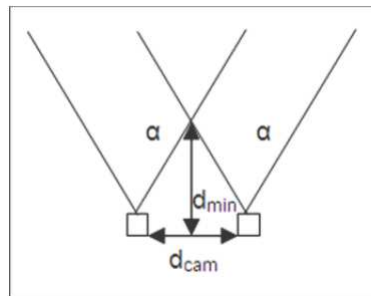


Rys. 5.4 Epipolarna geometria w układzie frontowo równoległym (kanonicznym)

Tak zorientowany zestaw kamer tworzy frontowo-równoległy epipolarny układ kamer (rys. 5.4), w którym ten sam rzeczywisty punkt sceny (P) rejestrowany z dwóch punktów o ogniskowych w punktach C<sub>l</sub> i C<sub>r</sub>, różni się tylko współrzędną poziomą (p<sub>l</sub> i p<sub>r</sub>), zaś współrzędna pionowa jest taka sama. Frontowo-równoległy układ kamer można zapewnić odpowiednią kalibracją kamer lub poprzez odpowiednią korektę obrazów uwzględniającą ich wzajemne przesunięcie i obrót.

Na bazie frontowo-równoległego układu kamer budowanie mapy głębi uzyskujemy poprzez dopasowanie (ang. *matching*) poszczególnych punktów obrazu zarejestrowanych przez dwa punkty widzenia (kamery). Porównanie różnic położenia odpowiadających sobie punktów pozwala oszacować odległość od układu kamer. Najczęściej porównuje się wzajemne położenie na dwóch obrazach punktów, krawędzi, regionów lub innych cech szczególnych obrazu.

Podejście do szacowania głębokości na podstawie układu frontowo-równoległego ma również swoje ograniczenia. Nie da się w ten sposób oszacować odległości obiektów znajdujących się zbyt blisko kamer, które nie zawierają się w obszarze widzenia (rys. 5.5).

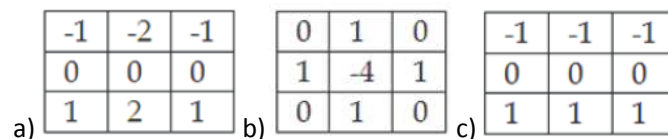


Rys. 5.5 Minimalna odległość  $d_{min}$ , poniżej której nie można szacować odległości obiektów sceny, będącą funkcją kąta widzenia  $\alpha$  i odległości pomiędzy kamerami  $d_{cam}$ .

Analogicznie można wprowadzić maksymalną odległość szacowania dysparycji, powyżej której punkty charakterystyczne nie są rozpoznawane jako sobie odpowiadające.

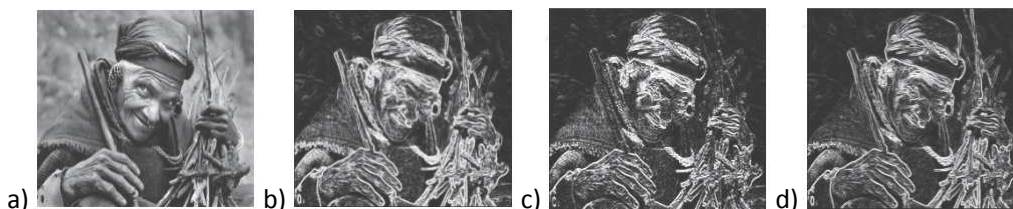
Jednym z podejść do szacowania głębi obrazu jest oparty o segmentację rejonów. Rejon to obszar, który zawiera zbiór podobnych pikseli. Obraz jest wtedy dzielony na zbiór rozłącznych, jednorodnych regionów, które konstruowane są na bazie określonej charakterystyki [Pha00, Gao06, Esp06, Wan08].

Kolejnym podejściem jest wzajemne lokalizowanie w obrazach charakterystycznych punktów i krawędzi. Najczęściej punkty i krawędzie charakterystyczne są znajdowane za pomocą operatorów przestrzennych. Przykładami takich operatorów są dyskretne filtry Sobela, Laplace'a czy Prewitt'a (rys. 5.6).



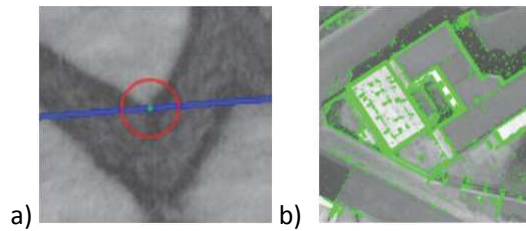
Rys. 5.6. a) detektor poziomych krawędzi Sobela; b) dyskretny operator Laplace'a; c) poziomy operator Prewitt'a

Na rysunku 5.7 zobrazowano wynik działania powyższych operatorów wykrywających krawędzie.



Rys. 5.7 a) rysunek oryginalny; b) dwukierunkowe filtrowanie Sobela; c) dwukierunkowe filtrowanie Prewitt'a; d) filtrowanie Laplasjanem [Ran05]

Podobnie rozpoznaje się punkty charakterystyczne w obrazach (rys. 5.8)



Rys. 5.8 Rozpoznawanie punktów charakterystycznych [Yu08] a) narożników na linii epipolarnej; b) zielono zaznaczone charakterystyczne punkty znalezione na całym obrazie

Algorytmy efektywnie obliczające mapę głębi na podstawie pary obrazów pochodzących z dwóch kamer zostały przedstawione w [Sza11]. Autorzy wykorzystują szybkie obliczanie rektyfikacji kamer, usuwanie zniekształceń obrazu i finalne obliczanie mapy dysparycji obrazów, której wynik koduje głębie obserwowanego obrazu.

Andrysiak i Choraś [And06] proponują klasyczne podejście do budowania mapy dysparycji na podstawie budowania relacji pomiędzy momentami obrazów, obliczonych dla dwóch punktów widzenia.

Bogate opracowanie odnośnie technik i algorytmów znajdowania punktów charakterystycznych w obrazach i ich dopasowywania można znaleźć w pracy Cyganka [Cyg09].

Problem przy konstruowaniu mapy głębokości na bazie obrazu pojawia się ze względu na głębię ostrości. Systemy optyczne kamer podobnie jak system wzrokowy człowieka charakteryzują się określoną głębią ostrości ustawioną zawsze względem punktu (obiektu), który znajduje się w obszarze koncentracji kamery/człowieka. Niestety poza obszarem głębi ostrości pojawia się nieostry obraz, którego stopień rozmycia może wskazywać zarówno na znajdowanie się przed jak i za strefą głębi ostrości.

Brak możliwości określenia głębi bezpośrednio w każdym pikselu obrazu wymusza konieczność interpolacji. Uzyskujemy wtedy interpolowaną mapę głębokości zamiast gęstej mapy głębokości. Innym pojęciem związanym z niedoskonałością konstruowania mapy głębokości jest jej względność albo bezwzględność. Najczęściej mamy do czynienia z względnymi mapami głębokości, w których jesteśmy w stanie jedynie określić który z pikseli znajduje się bliżej, a który dalej. Możliwość oszacowania bezwzględnej odległości od kamery do fragmentu sceny reprezentowanego przez konkretny piksel pozwala zbudować bezwzględną mapę głębokości.

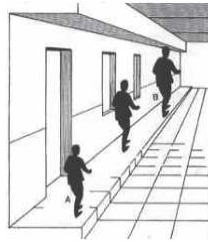
Podczas konstruowania mapy głębokości można wzorować się na systemie wzrokowym człowieka, który w sposób naturalny radzi sobie z szacowaniem głębokości obiektów znajdujących się w polu widzenia. Wśród mechanizmów wykorzystywanych przez ludzi do szacowania głębokości należą:

- para obrazów pochodzących z pary oczu. Niestety z mechanizm frontowo-równoległego ustawienia kierunków patrzenia mamy do czynienia tylko, jeśli obiekty znajdują się w nieskończoności. Dla obiektów bliższych kierunki patrzenia oczu są zbieżne a ludzki mózg wnioskuje o odległości punktu koncentracji m.in. na bazie kąta zbieżności i ewentualnej paralaksy obrazów;
- punkt koncentracji wzroku i odpowiadająca mu głębia ostrości to kolejny mechanizm, który człowiek może często zmieniać i na podstawie wyciągniętych wniosków mózg próbuje oszacować głębokość punktu/płaszczyzny koncentracji;
- dodatkowe cechy obrazu i jego elementów tj.: kształty obiektów [Kur04], ich powierzchnie [Mes09], kolory [Jac07], ruch [Str84], wzorce [Geo76], lub ich połączenie [Gut07] wspomagają proces szacowania odległości obiektów od obserwatora;
- różnice w jasności, które obserwowane są przy jednorodnym oświetleniu (np. zewnętrznym – rys. 5.9) również mogą być wskazówką przy szacowaniu odległości od obserwatora



Rys. 5.9 szacowanie głębokości sceny z mgłą.

- struktura obserwowanej sceny wskazuje na pewne informacje o głębokości poszczególnych obiektów, chociaż mózg może popełniać przy okazji błędy (rys. 5.10);

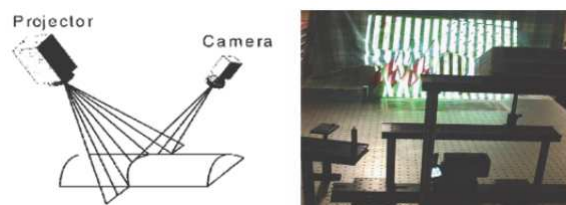


Rys. 5.10. Wizualna deformacja obiektów A i B względem postrzeganej struktury głębi sceny.

#### 4.1 Aktywne metody budowania mapy dysparycji

Aktywne metody szacowania głębokości polegają w głównej mierze na wysłaniu w głąb sceny pewnej dodatkowej energii, którą jest oświetlona scena, a następnie pasywnie analizuje się energię odbitą od obiektów. Konieczność wysłanie dodatkowej energii jest z jednej strony wadą tej grupy metod, ale charakteryzują się one dużo większą precyzją pomiarów.

Najczęściej dodatkowe wysłanie energii w głąb sceny jest realizowane poprzez dodatkowe oświetlenie sceny. O ile zwykle lampy żarowe nie dają zadowalającej precyzji, o tyle światło strukturyzowane, o znanym układzie wzorca, którym oświetlana jest scena, a następnie kamera z innej perspektywy ocenia zniekształcenia wzorca pozwala bardzo skutecznie ocenić głębokość poszczególnych elementów sceny.



Rys. 5.11. Schemat projekcji wzorca i oceny zniekształceń (potencjalnej głębokości) z perspektywy odpowiednio ustawionej kamery [Alb98].

Można również mierzyć czas, w jakim impuls światła przebył drogę od źródła do sensora, po drodze odbijając się od obiektu. Najczęściej w tego typu podejściu bazuje się na świetle podczerwym (IR) lub na świetle laserowym. Zaletą tego podejścia jest wysoka precyzja pomiaru (poniżej centymetra) oraz wysoka częstotliwość przetwarzania (100 fps). Na podobnej zasadzie wykorzystuje się ultradźwięki w popularnym badaniu USG, np. do badania płodu.

#### 4.2 Pasywne metody budowania mapy dysparycji

Metody nazywamy pasywnymi, jeśli bazują na naturalnym oświetleniu tła i na optycznej informacji zarejestrowanej na obrazach. Metody te rejestrują obrazy za pomocą sensorów obrazu i szacują mapę głębokości z pomocą obliczeń komputerowych. Wśród metod pasywnych można wyróżnić jeszcze dwie podgrupy metod: jedno-obrazowe (ang. *monocular*) i wiele-obrazowe (ang. *multiview*).

Projekt pt. *Osprzęt i oprogramowanie automatyzujące proces filmowania obrazu 3D* nr Umowy z NCBR: *INNOTECH-K3/H13/16/227587/NCBR/14* dofinansowany w ramach Programu „*INNOTECH*” w ścieżce programowej *HI-TECH*, współfinansowany ze środków Narodowego Centrum Badań i Rozwoju oraz Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego.

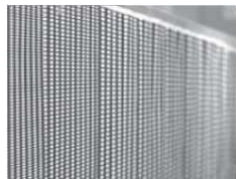
Metody jedno-obrazowe bazują na pojedynczej ramce lub sekwencji pojedynczych ramek obrazu. Bazują one głównie na obserwowanej strukturze obrazu i założonej strukturze obserwowanych obiektów wyposażenia [Fra01], znanych elementów ludzkiego ciała [Nag02] czy układu ścian i podłogi [Del05].

Można szacować również strukturę sceny na podstawie śledzenia ruchu punktów obiektów w czasie [Ozd07].



Rys. 5.12 Szacowanie odległości obiektów na podstawie wzajemnej odległości charakterystycznych punktów [Ozd07]

Powyższe rozwiązania pozwalają szacować odległość względną, ponieważ trudno jest na bazie jednego obrazu bezwzględnie oszacować odległości do obserwowanego obiektu. Jedyną metodą jedno-obrazową, która próbuje szacować bezwzględną odległość obiektu od kamery jest podejście bazujące na rozmyciu (ang. *defocusing level*) obrazu poza obszarem ostrości. Obliczenia przeprowadza się z wykorzystaniem operatora Laplace'a, który oblicza drugą pochodną przestrzenną każdego piksela na podstawie określonego jego sąsiedztwa [Hel01]. Porównanie stopnia rozmycia ze znanymi własnościami optycznymi kamery pozwala oszacować odległość, na jakiej znajdują się zarejestrowane fragmenty obrazu (rys. 5.13)



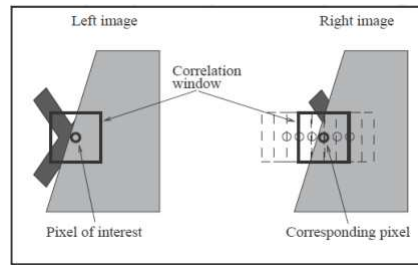
Rys. 5.13. Szacowanie odległości na bazie ostrości [Mal08].

Metoda ta jednak jest podatna na błędy szczególnie, jeśli stopień rozmycia obrazu jest znaczny lub, gdy pojawia się dodatkowy szum, na przykład spowodowany materiałem fotografowanego obiektu.

Drugą grupą metod pasywnych szacowania głębi sceny są metody oparte o dwa lub więcej obrazów rejestrujących scenę. Stereowizja jest szczególnym przypadkiem tych metod, gdy obraz jest rejestrowany przy pomocy dwóch kamer.

Kluczowym problemem przy metodach szacowania głębi opartych o dwa lub więcej obrazów jest problem dopasowania (ang. *matching*). Istnieją cztery główne nurty próbujące rozwiązać ten problem: lokalny, kooperacyjny, dynamicznego programowania i globalnej aproksymacji.

Pierwsza grupa bierze pod uwagę jedynie dysparycję wewnątrz skończonego, ograniczonego okna lub sąsiedztwa o podobnym rozkładzie jasności [Isl04, Wil98]. Porównuje się w tym celu wartości współczynnik dopasowania (SAD – suma wartości bezwzględnych różnic, SSD – suma kwadratów różnic lub dowolna inna charakterystyka sąsiedztwa piksela) dla każdej pozycji okna jest porównywana z innymi pozycjami. Okna są rozmiaru  $k \times k$  pikseli. Na dwóch obrazach znajduje się okna, których współczynnik dopasowania jest najbardziej podobny, a następnie dysparycja odciętych jest obliczana na podstawie pozycji dwóch odpowiadających sobie okien (rys. 5.14).



Rys. 5.14. Ruchome okno służące do odnalezienia przesunięcia krawędzi [Hir02]

Główną wadą tego podejścia jest liczba operacji, które muszą być przeprowadzone  $o(n) = N^3 \times k^4$ , gdzie  $N$  jest rozmiarem obrazu w pikselach, a  $k$  jest rozmiarem analizowanego okna. Rząd złożoności algorytmu powoduje, że nie są one najszybsze (1 do 5 fps [Hir02]).

Innym podejściem do lokalnego dopasowania polega na dopasowaniu punktów. Polega na odnalezieniu w obu obrazach zbioru punktów charakterystycznych, a następnie obliczenia ich dysparycji. Tego typu algorytmy też nie należą do najszybszych [Kim06, Liu06]. Dla przyjętego, bowiem rozkładu punktów charakterystycznych, tylko na bazie odnalezionych na różnych obrazach punktów mierzona jest dysparycja. Pozycję punktów niezidentyfikowanych należy interpolować, co powoduje błędy oszacowania. Metoda słabo radzi sobie z dużymi gładkimi powierzchniami, na których trudno zidentyfikować punkty charakterystyczne.

Drugą grupą jest podejście kooperacyjne, którego założeniem jest naśladowanie pracy ludzkiego mózgu. Sieć neuronowa iteruje zanim osiągnie stan stabilny [May03, Zit00].

Kolejnym podejściem jest podejście wzorowane na programowaniu dynamicznym. W podejściu tym porównywane są całe linie pomiędzy sobą. Podejście to bazuje na założeniu, że uporządkowanie linii w obrazie jest takie samo. Ewentualny problem dopasowania linii może być eliminowany poprzez łąpanie pewnych punktów odniesienia, np.: pionowych krawędzi [Oht85] lub poziomych podłogi [Bob99]. Podejście to należy do jednego z szybszych (50 fps [Kam08])

Grupa metod bazujących na globalnej optymalizacji poszukuje dysparycji, która minimalizuje globalny koszt lub funkcję energii dopasowania obszarów obrazów do siebie [Kac04].

Najprościej odnieść do obliczania mapy dysparycji dla układu kanonicznego. Najczęstszym przesunięciem obiektu w obrazie jest przesunięcie poziome, które oznacza dobrze skalibrowany rig stereoskopowy. Wtedy dysparycję liczymy ze wzoru:

$$D_x(p_L, p_R) = x_L - x_R = \frac{bf}{z}$$

gdzie  $x_L, x_R$  są odpowiadającymi sobie współrzędnymi tego samego punktu obiektu w obrazie lewym i prawym,  $b$  jest bazą układu kamer,  $f$  ogniskową, a  $z$  rzeczywistą odległością od obiektu do płaszczyzny bazowej (płaszczyzny matrycy).

W przypadku, gdy pojawia się dysparycja pionowa  $D_y$ , liczona analogicznie wzdłuż osi  $y$ , to wtedy sumaryczna dysparycja  $D$  jest obliczana z metryki euklidesowej.

$$D = \sqrt{D_x^2 + D_y^2}$$

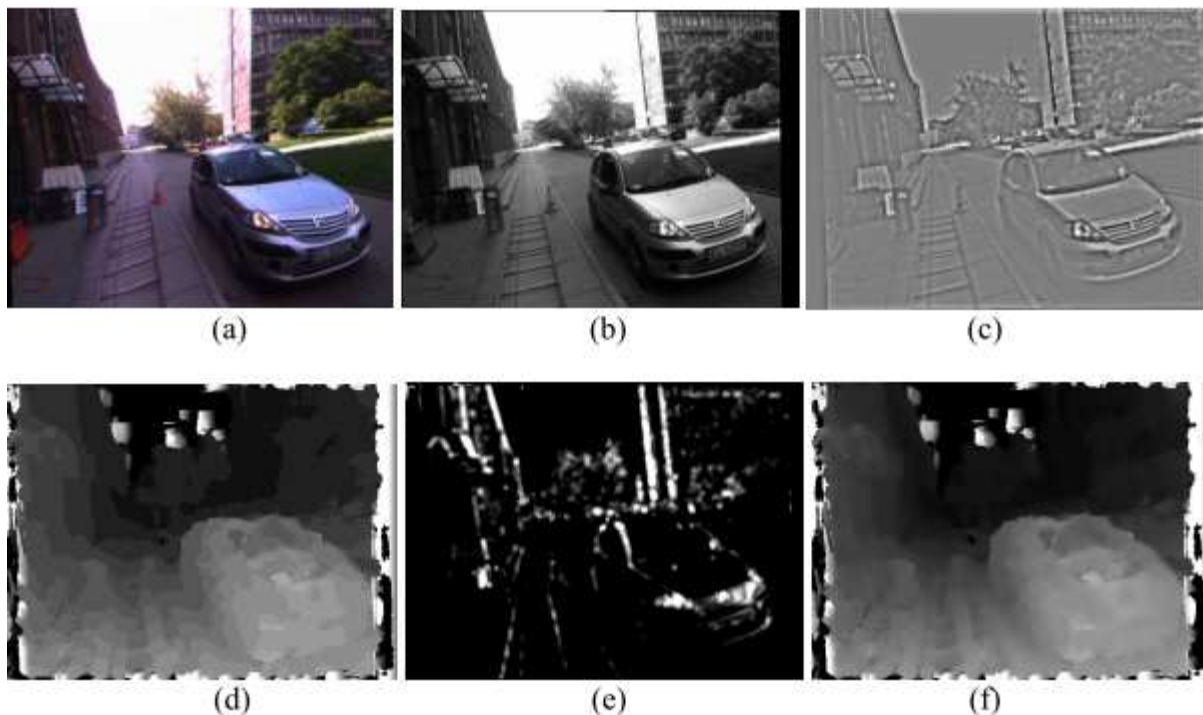
Zakres pomiaru głębokości obiektów sceny zależy dość istotnie od rozdzielczości matrycy i jej fizycznego rozmiaru. Pojedynczy piksel obrazu jest bowiem najmniejszą jednostką, o jaką może różnić się położenie obserwowanego punktu na dwóch obrazach. Dokładność szacowania głębokości zależy zatem, od fizycznej szerokości piksela ( $S_w/S_w$ ) i odległości, na której znajduje się obserwowany punkt ( $z$ )

$$\Delta z = \frac{\frac{S_w}{S_w} \cdot z^2}{f \cdot b - \frac{S_w}{S_w} \cdot z}$$

Przeprowadzając symulację na rzeczywistych parametrach dla bazy  $b$  rzędu 5 cm oraz rozdzielczości poziomej matrycy 1024 piksele i kącie widzenia  $60^\circ$  zakres skoku pomiaru głębokości na odległości 10 m wynosi nawet ok. 3m (dokładnie 2.91 m)

W rozwiązaniach mających na celu szybkie szacowanie mapy dysparycji celem określenia szacunkowej głębi poszczególnych elementów planu sugeruje się zastosowanie metod pasywnych opartych o analizę prostych cech, które mogą być szybko obliczane w kontrolerze riga dzięki wsparciu sprzętowemu (karty graficzne z GPU, układy FPGA). Przykładem takiej implementacji jest praca [Sza11], która oparta była o implementację na kartach graficznych i uzyskała bardzo dużą wydajność około 50 fps dla obrazów o rozmiarze 512x384 piksele. Charakterystyka układów przetwarzania równoległego wskazuje również, że dla większych rozmiarów obrazów odpowiednie zrównoleżenie przetwarzania obrazu na wiele układów przetwarzających równoległe nieznacznie obniży wydajność algorytmu.

Podstawowy algorytm polegał na próbkowaniu dla różnych wielkości okna  $N \times N$  (od  $3 \times 3$  do  $17 \times 17$ ) i różnych hipotetycznych wartości dysparycji  $d$  (od 1 do 9) wartości sumy wartości bezwzględnej różnic (SAD – ang. *sum of absolute differences*) koloru pikseli odpowiadających sobie okien w dwóch obrazach. Wartość dysparycji, dla której miara SAD była minimalna wskazywana była jako właściwa wartość dysparycji piksela będącego środkiem analizowanego okna. Takie obliczenie-chłonne podejście zostało 4-krotnie zoptymalizowane, co dało ostatecznie bardzo wydajne rozwiązanie (rys. 5.15).



Rys 5.15 Przykładowy zestaw obrazów w różnych etapach algorytmu [Sza11] a) obraz z prawej kamery; b) obraz przetransformowany do skali szarości c) obraz po filtrze górnoprzepustowym; d) mapa dysparycji otrzymana dla bloków  $15 \times 15$  pikseli; e) mapa pewności; f) podpikselowa mapa dysparycji;

## 5 Bibliografia

---

- [Alb98] Albrecht, P. and Michaelis, B. (1998). Improvement of the Spatial Resolution of an Optical 3D Measurement Procedure. *IEEE Transactions on Instrumentation and Measurement*, Vol.47, pp. 158-162.
- [Sch03] Scharstein D., Richard Szeliski. High-accuracy stereo depth maps using structured light. *Computer Vision and Pattern Recognition*, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 1. IEEE, 2003.
- [Kos03] Kostková, Jana, and Radim Sára. "Stratified Dense Matching for Stereopsis in Complex Scenes." *BMVC*. Vol. 5. 2003.
- [Sch10] Scharstein, Daniel, Baza danych par stereoskopowych i ich mapy głębi  
<http://vision.middlebury.edu/stereo/data/>
- [Sax08] Saxena, A., Chung, S. H., Ng, A. Y., 3-D Depth Reconstruction from a Single Still Image. *International Journal of Computer Vision*, Vol.76, pp. 53-69, 2008
- [Sza11] Szajerman D., Pełczyński P., Efficient implementation of stereovision algorithms for graphics processing unit in DirectShow technology, *Machine Graphics and Vision Int. Journal*, Vol. 20 issue 2, pp. 139-156, 2011
- [And06] Andrysiak T., Choraś M.: Algorithm for stereovision disparity calculation in the moment space, *Machine Graphics & Vision International Journal*, Vol. 15 Issue 3, February 2006.
- [Cyg09] Cyganek B.: *An Introduction to 3D Computer Vision Techniques and Algorithms*, John Wiley.
- [Pha00] Pham D.L., Xu C., Prince J.L. (2000). Current Methods in Medical Image Segmentation. *Annual Review of Biomedical Engineering*, Vol.2, pp. 315-337
- [Gao06] Gao L., Jiang J., Yang S.Y. (2006). Constrained Region-Growing and Edge Enhancement Towards Automated Semantic Video Object Segmentation. *Lecture Notes in Computer Science, Advanced Concepts for Intelligent Vision Systems*, Vol.4179, pp. 323-331.
- [Esp06] Espindola G.M., Camara G., Reis I.A., Bins L.S., Monteiro A.M. (2006). Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing*, Vol.27, pp. 3035-3040
- [Wan08] Wang, X. L. & Wang, L. J. (2008). Color image segmentation based on Bayesian framework and level set. *Proceeding of 2008 International Conference on Machine Learning and Cybernetics*, Vol.1, No.7, pp. 3484-3489
- [Ran05] Rangarajan, S. (2005), *Algorithms for Edge Detection*, Stony Brook University, <http://www.uweb.ucsb.edu/~shahnam/AfED.doc>
- [Yu08] Yu J., Weng L., Tian Y., Wang Y., Tai, X. (2008). A Novel Image Matching Method in Camera-calibrated System. In: *Cybernetics and Intelligent Systems*, 2008 IEEE Conference on, pp. 48-51
- [Kur04] Kurki I., Saarinen J. (2004). Shape perception in human vision: specialized detectors for concentric spatial structures?. *Neuroscience Letters*, Vol.360, pp. 100-102.
- [Mes09] Meese T. S., Summers R. J., (2009). Area summation in human vision at and above detection threshold. In: *Proceedings of the Royal Society B: Biological Sciences*, Vol.274, pp. 2891-2900.
- [Jac07] Jacobs G.H., Williams G.A., Cahill H., Nathans J. (2007). Emergence of Novel Color Vision in Mice Engineered to Express a Human Cone Photopigment. *Science*, Vol.315, pp. 1723-1727.



- [Str84] Stromeyer C.F., Kronauer R.E., Madsen J.C., et al. (1984). Opponent-Movement Mechanisms in Human-Vision. *Journal of the Optical Society of America A-Optics Image Science and Vision*, Vol.1, pp. 876-884.
- [Geo76] Georgeson M. (1976). Antagonism between Channels for Pattern and Movement in Human Vision. *Nature*, Vol.259, pp. 412-415.
- [Gut07] Guttman S., Gilroy L.A., Blake R. (2007). Spatial grouping in human vision: Temporal structure trumps temporal synchrony. *Vision Research*, Vol. 47, pp. 219-230.
- [Fra01] François A.R.J., Medioni G.G. (2001). Interactive 3D model extraction from a single image. *Image and Vision Computing*, Vol.19, pp. 317-328.
- [Nag02] Nagai T., Naruse T., Ikehara M., Kurematsu A. (2002). Hmm-based surface reconstruction from single images. In: *Image Processing. 2002. Proceedings. 2002 International Conference on*, Vol.2, pp. II-561 - II-564
- [Del05] Delage E., Lee H., Ng A. Y. (2005). Automatic single-image 3D reconstructions of indoor Manhattan world scenes. In: *12th International Symposium of Robotics Research (ISRR)*, pp.305-321.
- [Ozd07] Ozden K.E., Schindler K., van Gool L. (2007). Simultaneous Segmentation and 3D Reconstruction of Monocular Image Sequences. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1-8.
- [Hel01] Helmi F.S., Scherer S. (2001). Adaptive Shape from Focus with an Error Estimation in Light Microscopy. *2nd Int'l Symposium on Image and Signal Processing and Analysis*, pp. 188-193.
- [Mal08] Malik A.S., Choi T.-S. (2008). Depth Estimation by Finding Best Focused Points Using Line Fitting. *Lecture Notes in Computer Science, Image and Signal Processing*, Vol.5099, pp. 120-127.
- [Isl04] Islam M.S., Kitchen L. (2004). Nonlinear Similarity Based Image Matching. *International Federation for Information Processing*, Vol.228, pp. 401-410.
- [Wil98] Williams J., Bennamoun M. (1998). A Non-linear Filtering Approach to Image Matching. In: *Proceedings of the 14th International Conference on Pattern Recognition*, Vol.1, No.1, p. 3.
- [Hir02] Hirschmüller H., Innocent P.R., Garibaldi J. (2002). Real-Time Correlation-Based Stereo Vision with Reduced Border Errors. *Journal of Computer Vision*, Vol. 47, pp. 229-246.
- [Kim06] Kim H.K.I., Kogure K., Sohn K. (2006). A Real-Time 3D Modeling System Using Multiple Stereo Cameras for Free-Viewpoint Video Generation. *Lecture Notes in Computer Science, Image Analysis Recognition*, Vol.4142, pp. 237-249.
- [Liu06] Liu L., Gao H.-B., Zhang Q. (2006). Research of Correspondence Points Matching on Binocular Stereo Vision Measurement System Based on Wavelet. *CORD Conference Proceedings*, pp. 3687-3691. Available from: [http://pubget.com/paper/pgtmp\\_a32b66de88e2f5adb012c343cb5f2bf4](http://pubget.com/paper/pgtmp_a32b66de88e2f5adb012c343cb5f2bf4)
- [May03] Mayer H. (2003). Analysis of means to improve cooperative disparity estimation. *ISPRS Conference on Photogrammetric Image Analysis, ISPRS Archives*, Vol.XXXIV, Part 3/W8.
- [Zit00] Zitnick L., Kanade T. (2000). A cooperative algorithm for stereo matching and occlusion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.22, pp. 675-684
- [Oht85] Ohta Y., Kanade T. (1985). Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.7, pp. 139-154.
- [Bob99] Bobick A., Intille S. (1999). Large occlusion stereo. *International Journal of Computer Vision*, Vol. 33, pp. 181-200.

[Kam08] Kamiya S., Kanazawa Y. (2008). Accurate Image Matching in Scenes Including Repetitive Patterns. Lecture Notes in Computer Science, Robot Vision, Vol.4931, pp. 165-176.

[Kac04] Käck, J. (2004). Robust Stereo Correspondence using Graph Cuts. Master Thesis, Royal Institute of Technology. Available from:

[www.nada.kth.se/utbildning/grukth/exjobb/rapportlistor/2004/rapporter04/kackperjonny\\_04019.pdf](http://www.nada.kth.se/utbildning/grukth/exjobb/rapportlistor/2004/rapporter04/kackperjonny_04019.pdf).

Koniec