

Zdzisław KOWALCZUK, Tomasz MERTA

POLITECHNIKA GDAŃSKA, WYDZIAŁ ETI,
ul. Gabriela Narutowicza 11/12, 80-233 Gdańsk

Wizualizacja obrazu stereowizyjnego w systemie VISROBOT**Prof. dr hab. inż. Zdzisław KOWALCZUK**

Prof. dr hab. inż. (2003, 1993, 1986, 1978). Prof. zw. Automatyki i Robotyki oraz kier. katedry Systemów Decyzyjnych na wydziale ETI PG. Interesuje się sterowaniem adaptacyjnym, identyfikacją i estymacją, diagnostyką, detekcją, sztuczną inteligencją, robotyką, automatyką i informatyką. Kilkanaście książek (WNT, Springer, PWNT), około 100 artykułów oraz 300 referatów konferencyjnych i rozdziałów książkowych. Członek Komitetu Automatyki i Robotyki PAN, wiceprezes TKP (NOT) i prezes PolSPAR (IFAC).

e-mail: kova@pg.gda.pl

**Mgr inż. Tomasz MERTA**

Absolwent Wydziału Elektroniki, Telekomunikacji i Informatyki (ETI) Politechniki Gdańskiej. Obecnie pracuje jako wykładowca w Katedrze Systemów Decyzyjnych. Jest słuchaczem Studium Doktoranckiego WETI. Jego zainteresowania naukowe obejmują przetwarzanie obrazu, systemów wizyjnych w robotyce w tym zagadnienia stereowizji.

e-mail: tomasz.merta@pg.gda.pl

**Streszczenie**

Proponując nowe podejście do wizji robotycznej, opisuje się system, którego zadaniem jest wizualizacja obrazu stereo dla operatora robota. Pomysł ten zasada się na inteligentnym wykorzystaniu zmiennej bazy stereo, która pozwala na zwiększenie zakresu rozdzielczości głębi obrazu dla obiektów odległych. Na podstawie obrazu stereo, przy zmiennej bazie stereo, wyznaczana jest mapa dysparycji. Odpowiednio przeprowadzona wizualizacja (zakresu) dysparycji umożliwia poprawne postrzeganie głębi w obrazie stereoskopowym dostarczającym ludzkiemu operatorowi.

Słowa kluczowe: stereowizja, stereoskopia, SLAM.

Stereo image visualization in a VISROBOT system**Abstract**

The article describes a novel approach to robotic vision in mobile robotic systems. The task of such a system is to visualize a stereo image properly for an operator. The system uses different stereo baseline values. Variable baseline can result in increasing depth resolution for distant objects. We assume that the robot works in a static environment. Variable baseline stereo imaging is realized based on the submission of images taken from different robot's placements. A disparity map is determined with the use of various baselines. Visualization of the disparity extent allows making a stereoscopic picture that ensures a proper perception of the depth.

Keywords: stereovision, stereoscopy, SLAM.

1. Wprowadzenie

Głównym zadaniem system wizyjnego jednostek autonomicznych jest poznawanie otoczenia, w sposób podobny do zmysłu wzroku człowieka. W połączeniu z innymi zmysłami wzrok umożliwia naukę o otaczającym środowisku. Omawiany system pozwala na rozpoznawanie obiektów statycznych i ruchomych oraz na postrzeganie zależności między nimi.

W przypadku robotów mobilnych system wizyjny jest silnie ograniczony i przystosowany do konkretnego zadania. Możliwości algorytmów przetwarzających obraz obecnie nie pozwalają na analizę złożonych sytuacji. Choć same parametry kamery mogą być zbliżone do parametrów oka, to aktualnie stosowane algorytmy przetwarzania obrazu nie dają jeszcze zadowalających rezultatów w porównaniu z możliwościami człowieka.

Istnieją implementacje robotów mobilnych, które są zaopatrzone w system wizyjny umożliwiający realizację rozmaitych funkcji, np. mobilnego odkurzacza, pojazdu podwodnego, robota humanoidalnego, piłkarza, czy robota rozpoznającego teren.

W analizowanym kontekście głównym zadaniem systemu jest akwizycja i wizualizacja stereowizyjnego obrazu dla operatora robotów, a dokładniej - przekazanie obrazu stereoskopowego oddalonych obiektów bez zbliżania się do nich (na przykład w przypadku badania otoczenia skażonego).

Wykorzystując zmienną bazę stereo można zbudować obraz stereoskopowy odległych obiektów z odpowiednio dużą głębią.

2. SLAM

Oprócz realizacji nadrzędnego zadania robot mobilny musi bezpiecznie poruszać się w swoim otoczeniu, omijając przeszkody i poruszając się tak, aby nie ulec uszkodzeniu. Możliwe jest to w przypadku znanego i ograniczonego środowiska. Natomiast praca w nieznanym otoczeniu wymaga realizacji przez robota funkcji jednoczesnego określania własnej pozycji oraz mapowania (ang. *Simultaneous Localization and Mapping - SLAM*). Znane położenie robota oraz przeszkód w jego otoczeniu stanowi zwykle podstawę funkcjonowania robota.

Typowa realizacja SLAM opiera się na sensorach, które zbierają dane z najbliższego otoczenia robota. Obecne rozwiązania wykorzystują rozmaite czujniki, takie jak: lidar, sonar, dalmierz laserowy i kamery wizyjne [2, 3]. Dobiera się je w zależności od zastosowania, prędkości poruszania się robota oraz środowiska pracy. Implementacja funkcji SLAM jest niezbędna dla zautomatyzowanego procesu obserwowania otoczenia, bez nadzoru operatora.

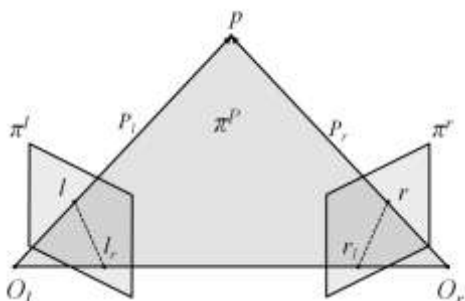
Funkcja SLAM rozpatrywana jest w przestrzeni dwu- lub trójwymiarowej (w zależności od potrzeb). Generowanie mapy 2D w celach nawigacyjnych jest wystarczające, gdy robot porusza się po terenie płaskim, gdzie przeszkodami są wysokie bryły (jak np. ściany). W bardziej złożonych przypadkach konieczne może być uwzględnienie szczegółowej mapy 3D.

Warto tu zaznaczyć, że kamery mają szczególne właściwości. Są sensorami pasywnymi (nie wysyłają aktywnie sygnałów w celu badania otoczenia) oraz dostarczają dane, które opisują obiekty (znajdujące się) w bliskim lub dalekim otoczeniu. Przy użyciu dwóch lub więcej kamer można uzyskać obraz stereowizyjny (podobnie, jak robi to człowiek). Także obraz z pojedynczej kamery znajdującej się w ruchu, pozwala na tworzenie map 3D - wykorzystując grupę technik SfM (ang. *Structure from Motion*). Ze względu na stosowanie tylko jednej kamery technikę tę określa się także terminami *MonoSLAM* lub *Single Camera SLAM*. W technice SfM odległość między kamerą a danym punktem obliczana jest na podstawie analizy różnic w położeniu (wybranych) punktów charakterystycznych (będących efektem przetwarzania obrazu) w kolejnych klatkach obrazu (ruchomej kamery) [4]. Stąd technikę SfM można stosować tylko dla obiektów znajdujących się w bliskim otoczeniu kamery. Należy także zapewnić odpowiednie ograniczenia ruchu kamery. Prezentowane podejście natomiast jest koncepcyjnie uproszczone. Stereowizja pozwala na wyznaczenie głębi w obrazie i wykrycie określonych obiektów oraz ułatwia dalszą analizę obrazu.

3. Stereowizja

Stereowizja naśladuje system widzenia człowieka, dlatego wizualizacja obrazu obserwowanego otoczenia na podstawie danych stereowizyjnych jest niezwykle użyteczna dla operatora. Dwie kamery (stereopara) – analogicznie jak para oczu – dostarczają obrazy, które przedstawiają obiekty z dwóch perspektyw. Zależności geometryczne pomiędzy punktem w przestrzeni i jego obrazem w obu kamerach są opisane za pomocą geometrii epipolarniej.

Na rys. 1 widoczne są lewy i prawy środek rzutowania (O_l, O_r) oraz płaszczyzny obrazów (π^l, π^r) odpowiednio lewej i prawej kamery. Płaszczyzna π^p przechodząca przez punkt p oraz oba środki rzutowania O_l, O_r tworzą płaszczyznę epipolarną. Jej przecięcie z płaszczyznami obrazów π^l, π^r na matrycy kamer tworzy sprzężone linie (odcinki) epipolarnie (l_l i l_r), zaś obraz r_l środka rzutowania O_l w prawej kamerze oraz obraz r_r środka rzutowania O_r w lewej kamerze stanowią środki epipolarnie. Odcinek łączący środki rzutowania O_l, O_r , jest bazą stereo (ang. *stereo baseline*).

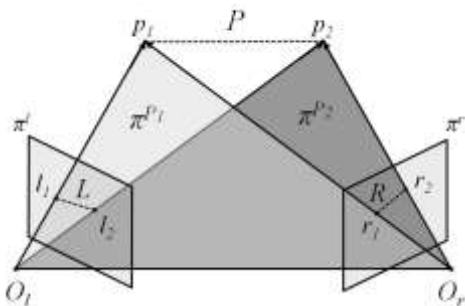


Rys. 1. Geometria epipolarna
Fig. 1. Epipolar geometry

Choć położenie i ułożenie kamer w przestrzeni jest nieznanne, dany punkt p w przestrzeni opisany jest wektorami (P_l, P_r) zaczepionymi w środkach rzutowania O_l i O_r w każdej z kamer. Powiązanie tych wektorów jest możliwe po kalibracji, tj. usunięciu zniekształceń geometrycznych kamer i wyznaczeniu parametrów zewnętrznych kamer. Parametry te opisują przekształcenie przestrzeni za pomocą wektora translacji T oraz rotacji R [5]. Relacja między wektorami P_l i P_r uzyskuje wtedy oczywistą postać:

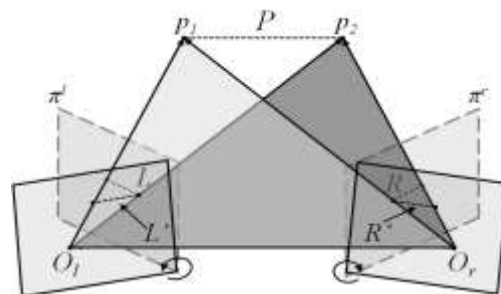
$$P_r = R(P_l - T) \quad (1)$$

Znajomość wektorów T i R jest konieczna do dalszej analizy, gdyż w ogólności szukanie rzutów punktu 3D na płaszczyznach π^l, π^r jest zadaniem trudnym. Dany odcinek rzeczywisty P ($p_1 p_2$) równoległy do układu kamer (bazy stereo $O_l O_r$) można rzutować na płaszczyzny π^l, π^r i otrzymać odpowiednio odcinki L ($l_1 l_2$) i R ($r_1 r_2$) widoczne na rys. 2.



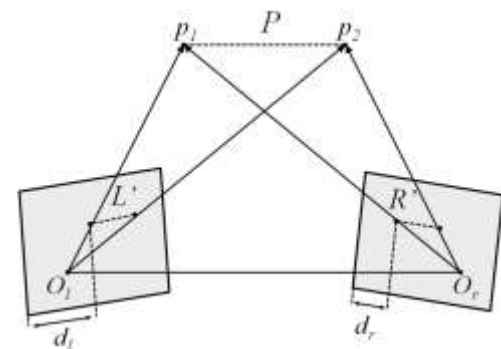
Rys. 2. Rzutowanie odcinka na płaszczyzny nieskalibrowanych kamer
Fig. 2. Projection of a segment on an uncalibrated camera planes

W ogólności odcinki te nie są względem siebie równoległe, co więcej nie pokrywają się z poziomymi liniami na płaszczyznach kamer (π^l, π^r). Dlatego dokonywana jest rektyfikacja obrazów, tj. takie przekształcenie punktów na płaszczyznach π^l, π^r , aby rzuty odcinka P znajdowały się w tej samej linii poziomej na matrycach π^l i π^r . Po poprawnej rektyfikacji sprzężone odcinki L i R uzyskują postać odcinków L' i R' , które mieszczą się na tej samej (np. 5 od góry) linii poziomej na płaszczyznach π^l, π^r (rys. 3).



Rys. 3. Rzutowanie odcinka na zrektyfikowane płaszczyzny kamer
Fig. 3. A segment projection on a rectified camera planes

Po rektyfikacji wyznacza się dysparycję, która określa w pikselach ($d_l - d_r$) przesunięcie obrazu danego punktu 3D na płaszczyznach π^l, π^r (rys. 4). Wartości ($d_l - d_r$) dla wszystkich punktów 3D są prezentowane w postaci mapy dysparycji. Uwzględniając dodatkowo parametry wewnętrzne kamer, punktom na mapie dysparycji można przypisać odległości w przestrzeni, tj. wyznaczyć mapę głębi.



Rys. 4. Rzutowanie odcinka na zrektyfikowane płaszczyzny kamer
Fig. 4. Projection of a segment on a rectified camera planes

Głębina w obrazie, czyli odległość danego punktu 3D od układu kamer (tj. bazy stereo), wyznaczana jest następująco [9]:

$$z = \frac{Bf}{d_l - d_r} = \frac{Bf}{d} \quad (2)$$

z – głębina obrazu, m
 B – szerokość bazy stereowizyjnej, m
 f – ogniskowa, piksele
 d – dysparycja, piksele,

zaś błąd wyznaczania głębiny obrazu Δz (w metrach) uwarunkowany jest dodatkowo błędem dysparycji Δd (w pikselach), zgodnie z przepisem:

$$\Delta z = \frac{z^2}{Bf} \Delta d \quad (3)$$

gdzie

$$f = \rho \cdot \Theta \quad (4)$$

Θ - rozmiar piksela (w metrach), (ogniskowa w metrach)

ρ - rozdzielczość kamery, $\frac{\text{piksel}}{\text{m}}$.

Zgodnie z (2) i (4) wartość z jest zależna od ustawień ogniskowej f oraz wymaga znajomości rozmiaru pojedynczego piksela Θ matrycy kamery w metrach (f w metrach). Im większa jest dysparycja, tym dany element znajduje się bliżej układu kamer. Błąd pomiaru odległości rośnie wraz z kwadratem odległości. Zachowanie stałej dokładności pomiaru wymaga liniowego zwiększenia jednocześnie B oraz f wraz z liniowym wzrostem z [6].

Na wartość f można wpłynąć bezpośrednio poprzez faktyczną jej zmianę, uzyskując np. zbliżenie (ang. *zoom*). Optyka z dużym zakresem zoomu, tj. dużym zakresem zmian ogniskowej f wymaga zastosowania kosztownego i dokładnego obiektywu. Ponadto duży zoom ogranicza pole widzenia kamery (ang. *field of view*) do środkowego obszaru obserwacji (środku matrycy kamery).

Zmianę zakresu f przeprowadzić można też zmieniając rozdzielczość obrazu. Zmianie ulegnie rozmiar pojedynczego piksela Θ i w konsekwencji f . Skoro rozdzielczość kamery ρ jest ustalona, rozdzielczość obrazu można jedynie zmniejszyć dla małych odległości z (co zmniejsza zapotrzebowanie na moc obliczeniową przy dalszym przetwarzaniu obrazu). Zmniejszenie rozdzielczości kamery, w konsekwencji także zwiększenie błędu wyznaczania głębi w obrazie, jest istotne, gdy priorytetem jest zastosowanie stałej mocy obliczeniowej przy różnych wartościach z [6]. Z kolei próba modyfikacji dokładności dysparycji d , na przykład poprzez skorzystanie z wartości subpikselowych przy wyznaczaniu d , jest nieskuteczna ze względu na utożsamianie (ang. *aliasing*) [7].

Poszerzenie bazy stereo B można uzyskać poprzez oddalenie od siebie kamer. Dla robota mobilnego z przesuwanymi kamerami dostępny zakres zmian B może być niewystarczający. Przy założeniu pracy w statycznym otoczeniu, efekt zmiany bazy można uzyskać poprzez przemieszczanie się robota. Parą obrazów stereo w tym przypadku jest obraz z kamery z poprzedniego położenia robota oraz obraz z kamery z aktualnego położenia robota. Dla grupy robotów, taką zmienną bazę można uzyskać rozdzielając stereoparę pomiędzy 2 wybrane roboty.

4. Wizualizacja obrazu z kamer robota

Obraz 3D można przedstawić operatorowi (wyświetlając odpowiednio lewy oraz prawy obraz), jeśli spełnione są ograniczenia wynikające z własności systemu wizyjnego człowieka. Korzysta się tu z okularów i techniki anaglifowej, polaryzacyjnej lub migawkowej, albo z wyświetlaczy autostereoskopowych, które nie wymagają okularów. Jednak ze względu na specyfikę obrazu generowanego przez oko ludzkie oraz ograniczenia wyświetlaczy 3D, obraz z kamer stereo musi być tak dostosowany, aby zapewnić prawidłową fuzję stereopary w mózgu obserwatora [8].

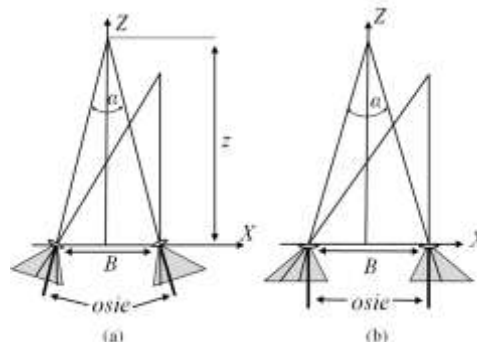
4.1. Wyświetlenie obrazu stereoskopowego

Obrazy z kamer o określonej konfiguracji różnią się od obrazów generowanych przez oczy obserwatora, dlatego inny jest sposób wyznaczania mapy głębi w obu przypadkach. Wzrok 'wpatrzony' jest w jeden dowolny punkt, skutkiem czego osie optyczne oczu spotykają się w tym punkcie (rys. 5a). Dla systemu stereowizyjnego konfiguracja ta jest niepraktyczna. Zwykle stosuje się kanoniczny (równoległy) układ kamer (rys. 5b), gdzie osie optyczne kamer są do siebie równoległe, co znacznie upraszcza porównywanie obrazów przez algorytm stereowizyjny.

Dodatkowo, jeśli ważne jest, aby człowiek widział zobrazowane otoczenie dokładnie tak, jakby się znajdował w miejscu robota, należy usunąć zniekształcenia stereoskopowe. Operacja ta sprowadza się do zapewnienia, aby bazy stereo kamer oraz oczu były

równe, pole widzenia obserwatora było takie, jak pole widzenia kamery, a także aby nie występowała zmiana rozdzielczości głębi (ang. *magnification of depth*) [10]. Zwykle jednak priorytetem nie jest sprawienie wrażenia, że człowiek znajduje się w miejscu robota, ale obserwowanie otoczenia i przybliżanie istotnych detali.

Do poprawnej syntezy obrazu stereoskopowego potrzebny jest opis zależności głębi obrazu od charakterystyki wyświetlacza oraz warunków wyświetlania.



Rys. 5. Układ kamer: (a) wpatrzony; (b) równoległy
Fig. 5. Camera system: (a) toad-in; (b) parallel

Dysparycja d opisuje różnicę między obrazami (na matrycach kamer), zaś paralaksa P opisuje różnicę między obrazami na wyświetlaczu 3D. Paralaksa występuje, gdy obserwacja obrazu na wyświetlaczu 3D jest realizowana za pomocą układu wpatzonego (rys. 6). Można zatem w wybranym przypadku utożsamiać te dwa pojęcia ze sobą. Okazuje się, że jeśli obraz z matryc (L i R) na rys. 2 zostanie poddany procesowi rektyfikacji i skalowania adekwatnie do wymiarów wyświetlacza to można utożsamiać paralaksę P z dysparcją d :

$$P = \frac{W_D}{W_S} d \quad (5)$$

W_D – szerokość wyświetlacza,

W_S – szerokość matrycy kamery,

d – dysparycja.

Paralaksa może przyjmować wartości dodanie (rys. 6), co ma miejsce w przypadku, gdy obserwator ma wrażenie, że punkt 3D znajduje się za ekranem (wzrok jest skupiony na punkcie, który znajduje się za ekranem). Wartości ujemne paralaksy występują, gdy wzrok jest skupiony na punkcie, który znajduje się przed ekranem. Zerowa paralaksa daje naturalny obraz.

Korzystając ze znajomości odległości obserwatora od ekranu (rys. 6), szerokości bazy stereo obserwatora oraz paralaksy w ogólności możliwa jest rekonstrukcja głębi (twierdzenie Tale-
sa):

$$z_v = \frac{z_D B_h}{B_h - P} \quad (6)$$

z_v – głębia w obrazie, m,

z_D – odległość ekranu od obserwatora, m,

B_h – baza stereo narządu wzroku (średnio 0,064 m),

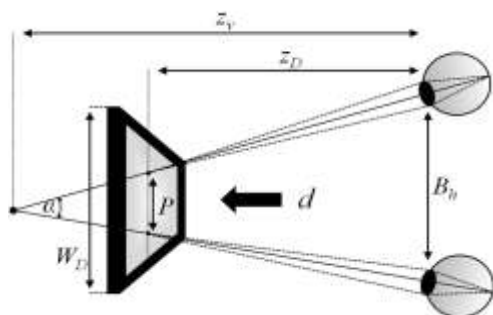
P – paralaksa, m.

Zatem korzystając z (5) i (6) głębię odczuwaną przez obserwatora ludzkiego wyrazić można w funkcji dysparycji:

$$z_v = \frac{z_D B_h}{B_h - \frac{W_D}{W_S} d} \quad (7)$$

4.2. Komfortowy zakres widzenia

Układ wpatrzony narządu wzroku dodatkowo charakteryzuje się tym, że kształt soczewki oka zmienia się zależnie od obiektu, na którym wzrok jest skierowany i skupiony. Dla dowolnego rzeczywistego punktu 3D narząd wzroku dostosowuje ogniskową i kąt patrzenia. Akomodacja narządu wzroku dostosowuje ogniskową, aby obiekt w obrazie był ostry, zaś konwergencja (ang. *convergence*) – ustawia kąty patrzenia każdego z oczu tak, aby obiekt ustawiony był na środku obu obrazów stereo (rys. 6). Jeżeli jednak punkt ten znajduje się w ekranie, oko naturalnie dostosowuje ogniskową do odległości z_D obserwatora od ekranu, zaś kąt patrzenia (konwergencji) α dostosowywany jest do odległości z_v , jeśli obserwowany obiekt znajduje się za ekranem lub przed ekranem (rys. 6). Z powodu konfliktu akomodacji i konwergencji różnica między z_D i z_v nie może być duża.



Rys. 6. Postrzeganie głębi w wyświetlanym obrazie stereoskopowym
Fig. 6. Perception of depth in a displayed stereoscopic image

Na drodze empirycznej wyznaczono zakres zmian kąta konwergencji α , który nie zakłóca poprawnej fuzji obrazu stereoskopowego w mózgu człowieka [9]. Różnica między wartością maksymalną i minimalną (tego zakresu) wynosi $\Delta\alpha = 0,02$ rad. Znajac $\Delta\alpha$ można obliczyć maksymalny zakres paralaksy – zgodnie z (5) opisanym w sekcji 4.1 – i odpowiadający jej zakres dysparycji. Stosowanie jednej wartości $\Delta\alpha$ do całego obrazu nie jest odpowiednie. Co więcej tolerancja na większe wartości parametru $\Delta\alpha$ zależy od indywidualnych cech obserwatora.

Ustalając zakres konwergencji $\Delta\alpha$ oraz odległość od ekranu z_D można określić dopuszczalny zakres zmian paralaksy:

$$\Delta P = P_{\max} - P_{\min} = z_D \Delta\alpha \quad (8)$$

Odnosząc zakres paralaksy do rozmiaru ekranu uzyskujemy wartość względną:

$$\Delta P_{rel} = \frac{\Delta P}{W_D} = \frac{z_D}{W_D} \Delta\alpha \quad (9)$$

Jak widać, względny zakres paralaksy zależy od stosunku odległości obserwatora (od ekranu) do szerokości ekranu. Źródła [9] podają różne wartości optymalnej odległości od ekranu, większość z nich zawiera się w przedziale $\langle 1,2, 2,3 \rangle$. Przyjmując, że $z_D/W_D \approx 1,67$ [9] względna wartość zakresu paralaksy $\Delta P_{rel} \approx 1/30$. Jest to tzw. „reguła 1/30”, według której wartość paralaksy (dodatniej lub ujemnej) nie powinna być większa niż 1/30 szerokości ekranu. Przyjmując $P_{\max} = 1/30 W_D$, zgodnie z (5), uzyskujemy dopuszczalny zakres dysparycji $d_{\max} = W_D/30$ (w pikselach).

Bez uwzględniania innych efektów, które także przyczyniają się do pogorszenia stereoskopii [9], można stwierdzić, że zmiany parametrów kamery nie pogarszają znacznie obrazu stereo, pod warunkiem zachowania dopuszczalnego zakresu dysparycji.

5. System VISROBOT ze zmienną bazą

Robot z systemem stereowizyjnym, o znanej szerokość bazy może oceniać odległość od obiektu z większą dokładnością. Ponadto, przy zachowaniu ograniczeń na dopuszczalny zakres dysparycji (sekcja 4), za pomocą wyświetlacza 3D lub odpowiednich okularów można wyświetlić operatorowi poprawny obraz stereoskopowy. Szerokość bazy ograniczona jest przez maksymalną wartość dysparycji oraz konfigurację i ustawienie optyki kamer. Przy zachowaniu opisanych zależności, szeroka baza pozwala na większą rozdzielczość głębi obrazu. Dla obiektów oddalonych od układu kamer występuje wrażenie bliskości i ‘pogłębienia/zoomu przestrzennego’. Pozwala to na lepsze rozróżnienie odległości między obrazowanymi obiektami.

Podejście ze zmienną bazą zostało użyte np. w [1], jednak celem autorów było wyznaczenie dokładnej mapy głębi. Rozwiązanie to nie odnosi się do poprawnego wyświetlania obrazu 3D i zastosowań w robotyce mobilnej.

5.1. Strojenie bazy stereo

Na podstawie (2), zakładając dopuszczalną wartość dysparycji, bazę B można uzależnić od nieznannej odległości od kamery:

$$B = \frac{d_{\max}}{f} z \quad (10)$$

Przy stałej wartości ogniskowej f można zwiększać szerokość bazy B wraz ze wzrostem odległości obiektu od kamery z . Aby określić dopuszczalną B konieczna jest zatem znajomość przybliżonej odległości z obserwowanych obiektów. W tym celu można najpierw wyznaczyć z według (2) przy małej szerokości bazy B_0 , a następnie poszerzać bazę, aby poprawić rozdzielczość głębi dla obiektów położonych dalej. Należy jednak przestrzeżać, aby przy nowej szerokości bazy dysparycja nie przekroczyła założonej wartości d_{\max} . Obiekty o odległości z od kamery (o stałej dysparycji d) można oznaczyć jako grupy pikseli (w obrazie).

Należy liczyć się z możliwością wystąpienia w obrazie obszarów o nieznannej wartości z , które są wynikiem błędów algorytmu stereowizyjnego. Ponadto przeszkody mogą uniemożliwić przeszczenie się robota w wymagane miejsce. Może też wystąpić okluzja obiektów znajdujących się daleko przez obiekty znajdujące się blisko. Zatem obszary w obrazie, dla których nie udało się wyznaczyć dysparycji należy odrzucić oraz wykrywać okluzję przy zwiększaniu szerokości bazy. W przypadku systemu nadzorowanego przez człowieka problemy tego rodzaju można pozostawić do rozwiązania operatorowi.

5.2. Algorytm wizualizacji stereoskopowej otoczenia WSO w systemie VISROBOT

Zakładamy, że otoczenie robota jest nieruchome. Stereopara jest wizualizowana dla operatora w postaci obrazu 3D z możliwością uzyskania ‘pogłębienia przestrzennego’ (zwiększającego rozdzielczość głębi). Różne B realizowane są poprzez przemieszczenie kamer i złożenie obrazów z poprzedniego i aktualnego położenia robota (opisane w sekcji 1). Stąd konieczny jest dodatkowy sensor służący do wyznaczenia aktualnego położenia robota. Sensor ten musi podawać nie tylko aktualne położenie kamery, ale także kąty elewacji i azymutu kamery. Można tu wykorzystać: zintegrowany sensor (3-osiowy akcelerometr, grawimetr i magnetometr), odometrię, lub sygnał GPS (jeśli jest dostępny). Rozwiązanie zakłada użycie sensora, jednak problem dokładności wyznaczenia położenia w oparciu o zintegrowany sensor wymaga odrębnych badań.

Testowa implementacja opisanego systemu VISROBOT realizowana została na własnej konstrukcji robota mobilnego. Robot ten składa się z podwozia, kół, platformy PandaboardES, opartej na dwurdzeniowym procesorze ARM, sensora MinIMU-9 firmy Pololu (3-osiowy akcelerometr, grawimetr, magnetometr). System wizyjny robota zawiera 2 kamery USB IDS16-40LE z szerokokątnym obiektywem Eneo MegaPixel o ogniskowej 2,6 mm.

Algorytm WSO na początku generuje mapę głębi dla ustalonego (małego) B . (w testach przyjęto $B_0=10$ cm). Obraz stereo jest analizowany w celu wyznaczenia konturów. W pewnym przybliżeniu każdy wyznaczony kontur zawiera piksele o takiej samej wartości dysparycji (jednakowej odległości od kamer z). Dla wstępnie przetworzonego obrazu operator robota zaznacza obszar zainteresowania (kontur o danym $d=d_0$), który chce obserwować z większą rozdzielczością głębi. Po wskazaniu takiego konturu szerokość bazy stereoskopowej powinna być zwiększona.

Pojawia się tu konieczność ustalenia adekwatnych warunków obserwacji. Poza racjonalnym przypadkiem, gdzie dla zaznaczonego konturu dysparycja jest ograniczona $1 < d < d_{\max}$, możliwa jest graniczna sytuacja ('czarny punkt' na mapie dysparycji), gdzie $z=\infty$ i $d=0$ (gdyż $zd=Bf=const$). Jeśli dysparycja jest już mniejsza niż 1 piksel ($d_0 \approx 0$), ani nie można ocenić odległości z , ani nie można dobrać nowej bazy stereo B , przy której nowe d nie przekroczy wartości d_{\max} .

Zakładając jednak, że ów 'czarny punkt' ma jakąś określoną odległość z , którą określimy hipotetycznie korzystając z poprzedniej bazy B , uzyskujemy przybliżoną relację:

$$B_- \approx \frac{1 \cdot z}{f} \text{ lub } z = fB_- \quad (11)$$

Na podstawie powyższego przybliżenia oraz wzoru (10), przy założonym d_{\max} obliczamy maksymalną bazę stereo

$$B_{\max} \approx \frac{d_{\max}}{f} z = d_{\max} \cdot B_- \quad (12)$$

W ten sposób w ogólnym przypadku przestrzegamy zasady:

$$B_{\max} = \begin{cases} \frac{d_{\max}}{f} z \Leftrightarrow d_0 > 1 \\ B_- \cdot d_{\max} \Leftrightarrow d_0 \leq 1 \end{cases} \quad (12)$$

Ze względu na możliwe niedokładności algorytmu wyznaczania mapy dysparycji, nie należy zwiększać szerokości bazy w jednym kroku z $B=B_0$ do $B=B_{\max}$, ale zwiększać ją w kolejnych iteracjach algorytmu, np. o $\Delta B=B_0$.

Po obliczeniu nowej (zwiększonej) szerokości bazy stereo, robot przemieszcza się o wymaganą wartość w określoną, lewą albo prawą stronę. Następnie wyznaczane są parametry zewnętrzne kamery w oparciu o aktualne dane z sensora określającego położenie i ułożenie robota. Na podstawie uzyskanego obrazu oraz obrazu z kroku poprzedniego algorytm WSO oblicza aktualną odległość z dla konturu.

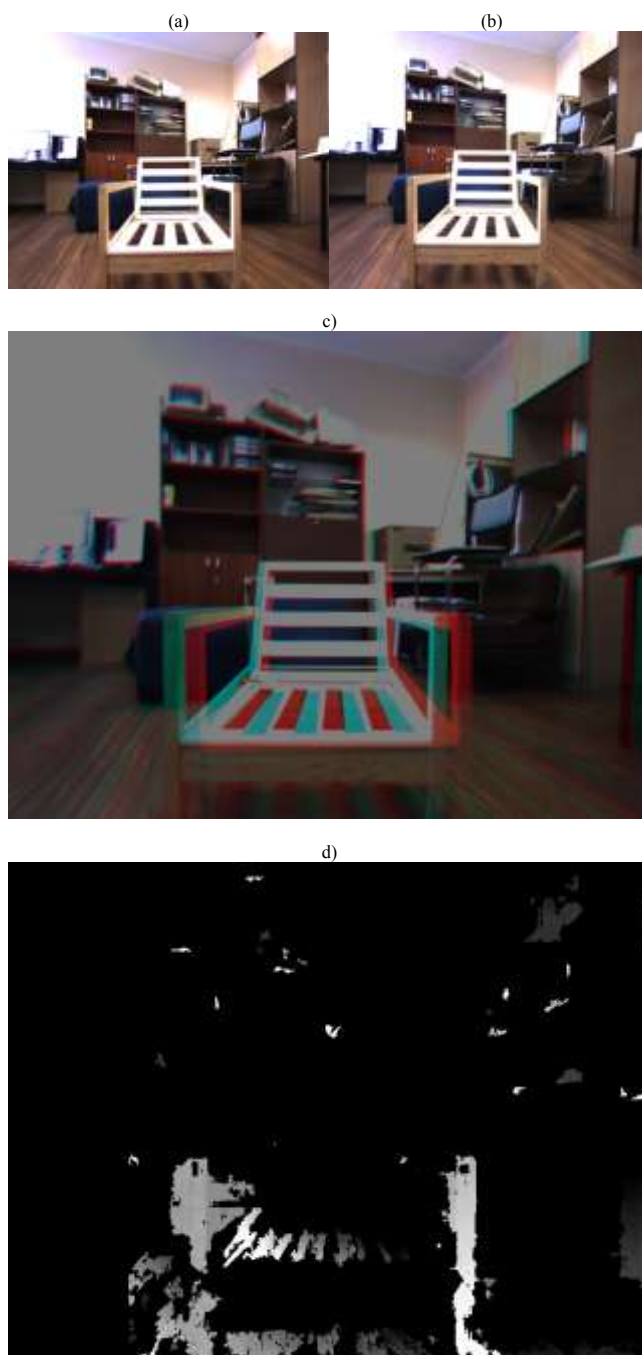
Algorytm kończy pracę, kiedy kontur zaznaczony przez operatora ma maksymalną dysparycję $d=d_{\max}$. Ponadto przed wyświetleniem obrazu stereoskopowego sprawdza się, czy w obrazie nie występują kontury, dla których $d > d_{\max}$. Obszary w obrazie o zbyt dużej dysparycji nie są wyświetlane operatorowi.

Procedura 1. Algorytm WSO

- Ustaw wartość początkową szerokości bazy $B=B_0$,
- generuj mapę dysparycji dla aktualnej B ,
- wyznacz kontury dla obszarów jednakowej dysparycji,
- oblicz z dla wszystkich wyznaczonych konturów,
- czekaj na wskazanie konturu (przez operatora),
- ustal B_{\max} ,

- zwiększ bazę stereo o ΔB ,
- wyznacz mapę dysparycji oraz d dla zaznaczonego obszaru,
- sprawdź, czy $d=d_{\max}$,
- jeśli tak, to ukryj kontury o $d > d_{\max}$ i zakończ,
- wykonaj transversalne przemieszczenie robota o zadany krok,
- odczytaj aktualne położenie i ułożenie kamer z sensora,
- zaktualizuj parametry zewnętrzne kamery.

Poprawność działania algorytmu WSO (Procedura 1), przy ograniczeniach opisanych w sekcji 4, można sprawdzić oglądając w okularach anaglifowych pełnokolorowe obrazy 3D [11]. Elementy w obrazie znajdujące się blisko są dobrze postrzegane w przypadku małej bazy (rys. 7c) i źle postrzegane w przypadku większej bazy (rys. 8c). Można także zauważyć, iż odległości między tłem a pierwszym planem są większe przy większej bazie stereo (większa rozdzielczość głębi).



Rys. 7. Zobrazowanie obiektu dla bazy stereo $B=7$ cm: a) obraz lewy, b) prawy obraz, c) dysparycja – obraz anaglifowy, d) 'bliska' mapa głębi

Fig. 7. Visualization of an object with baseline $B=7$ cm: a) left image, b) right image, c) disparity – anaglyph image d) 'close' depth map

Dla lepszego zobrazowania wyników przedstawiono także mapy dysparycji dla obu przypadków, gdzie nawet bez używania okularów można zaobserwować, jak zmienia się rozdzielczość głębi obrazu przy różnych bazach stereoskopowych. Czarny kolor oznacza, że odległość punktu od układu kamer jest nieznaną. Im kolor jest jaśniejszy (większa luminancja), tym dany punkt znajduje się bliżej układu kamer.

Rys. 7 oprócz rzeczywistego zdjęcia obiektu – obrazu lewego (a) i prawego (b), dysparycji (c), przedstawia ‘bliską’ mapę głębi (d) przy bazie stereo $B=7$ cm. Analogicznie, na rys. 8 zobrazowano ten sam obiekt przy $B=20$ cm.



Rys. 8. Zobrazowanie obiektu dla bazy stereo $B=20$ cm: a) obraz lewy, b) prawy obraz, c) dysparycja – obraz anaglifowy, d) ‘daleka’ mapa głębi
Fig. 8. Visualization of an object with baseline $B=20$ cm: a) left image, b) right image, c) disparity – anaglyph image d) ‘distant’ depth map

Porównując rys. 7d i 8d można zauważyć, że przy małej bazie stereo rozdzielczość głębi jest mała, (dla poprawienia czytelności

obrazu jasność oraz kontrast obrazu zostały zmodyfikowane, dlatego głębia na rys 7d ma inną skalę niż głębia na rys. 8d).

Porównując luminancje na rys. 7d i 8d można łatwo zaobserwować, iż obiekty bliskie są widoczne przy małej bazie, zaś duża baza lepiej odwzorowuje elementy znajdujące się daleko (co świadczy o większej rozdzielczości głębi). Warto zauważyć, iż na rys. 8d elementy, które mają bardzo dużą dysparycję są usunięte (obraz 8c dla obiektów bliskich nie jest poprawny). Problem ten uwzględnia się w Procedurze 1, która pozwala obserwatorowi na ingerencję umożliwiającą poprawną percepcję obrazu stereoskopowego.

6. Podsumowanie

W pracy przedstawiono system stereowizyjny wykorzystujący zmienną bazę stereo, odpowiedni dla zespołów robotów mobilnych. Dzięki technice stereowizyjnej robot (z dużej odległości) może pasywnie wyznaczać odległości obiektów znajdujących się w bliskim oraz dalszym otoczeniu. W szczególności, poprzez zmianę szerokości bazy stereo można w ten sposób dokładnie określać odległość obiektów znajdujących się na dalszym planie. W celu oceny takich oddalonych obiektów system zwiększa rozdzielczość głębi, wykorzystując efekt tzw. ‘pogłębienia przestrzennego’. Użyteczny obraz stereoskopowy może być na bieżąco wizualizowany i prezentowany operatorowi systemu.

Planowane jest zautomatyzowanie opisanego procesu mapowania przestrzennego z wykorzystaniem zmiennej bazy stereo z możliwością bieżącej wizualizacji użytecznej dla operatora ludzkiego. W kolejnym etapie badań warto rozważyć rozwój zespołu obejmującego współdziałające roboty z systemem stereowizyjnym, którego zadaniem jest praktyczne badanie statycznych elementów otoczenia.

7. Literatura

- [1] Okutomi M., Kanade T. (1993): A multiple-baseline stereo, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 15, no. 4.
- [2] Jinwoo Choi, Sunghwan Ahn, Minyong Choi, Wan Kyun Chung (2006): Metric SLAM in home environment with visual objects and sonar features, IEEE Transactions on Intelligent Robots and Systems, ss.4048-4053.
- [3] Newman P., Cole D., Ho K. (2006): Outdoor SLAM using visual appearance and laser ranging, Proceedings of 2006 IEEE International Conference on ICRA, ss.1180-1187.
- [4] Davison A.J., Reidet I.D., et al. (2007): MonoSLAM: Real-time single camera SLAM. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, ss. 1052-1067.
- [5] Kowalczyk Z., Węsierski D. (2007): Three-dimensional robot positioning system with stereo vision guidance, Proceedings of 13th IEEE/IFAC International Conference on Methods and Models in Automation and Robotics, ss. 1011-1016.
- [6] Gallup D., Frahm J.M., Mordohai P., Pollefeys M. (2008): Variable baseline/resolution stereo. IEEE Conference on Computer Vision and Pattern Recognition, ss. 1-8.
- [7] Szeliski R., Scharstein D. (2004): Sampling the disparity space image, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 3, ss. 419-425.
- [8] Siegel M., Nagata S. (2000): Just enough reality: comfortable 3-D viewing via microstereopsis, IEEE Transactions on Circuits and Systems for Video Technology, vol.10, no.3, ss.387-396.
- [9] Zilly F, Kluger J., Knauff P. (2011): Production rules for stereo acquisition. Proceedings of the IEEE, vol. 99, no. 4, ss. 590-606.
- [10] Jones G., Lee D., Holliman N., Ezra D. (2001): Controlling perceived depth in stereoscopic images, Proceedings of 8th Stereoscopic Displays and Virtual Reality Systems, ss. 42-53.
- [11] Kowalczyk Z., Merta T. (2013): Wizualizacja stereoskopowa obrazu, KSD, http://www.eti.pg.gda.pl/katedry/ksd/paper_online/index.html.