

Michał NOWICKI

POLITECHNIKA POZNAŃSKA, INSTYTUT AUTOMATYKI I INŻYNIERII INFORMATYCZNEJ,
Plac Marii Skłodowskiej-Curie 5, 60-965 Poznań

Porównanie metod opartych na cechach 2D i 3D w zadaniu samolokalizacji robota mobilnego na podstawie danych RGB-D

Inż. Michał NOWICKI

Inż. Michał Nowicki jest studentem studiów stacjonarnych II stopnia na kierunku „automatyka i robotyka” na Wydziale Elektrycznym Politechniki Poznańskiej oraz studiów stacjonarnych I stopnia na kierunku „informatyka”. Jest także laureatem konkursów MNiSW „Diamantowy Grant” oraz „Generacja Przyszłości”.



e-mail: michal.nowicki@put.poznan.pl

Streszczenie

Praca porusza problem estymacji transformacji przestrzennej pomiędzy kolejnymi pozycjami robota mobilnego. Transformacja wyznaczana jest na podstawie danych RGB-D uzyskanych z sensora Kinect w kolejnych pozycjach wzdłuż trajektorii ruchu robota. Bardzo duża liczba punktów w przestrzeni otrzymanych z sensora Kinect powoduje, że nie mogą być one przetwarzane w czasie rzeczywistym na większości komputerów pokładowych robotów mobilnych. Dlatego w pracy porównano dwie metody ekstrakcji punktów charakterystycznych (kluczowych) redukujących rozmiar zbioru danych: cechy oparte na detekcji w chmurze punktów 3D oraz cechy oparte na detekcji na dwuwymiarowym obrazie RGB. Metody te porównano pod względem czasu działania, skuteczności ekstrakcji punktów oraz dryftu estymaty pozycji robota na publicznie dostępnych zbiorach danych RGB-D.

Słowa kluczowe: robot mobilny, sensor Kinect, detektory cech, deskryptory cech, samolokalizacja.

Comparing robot self-localization methods exploiting 2D and 3D features in RGB-D data

Abstract

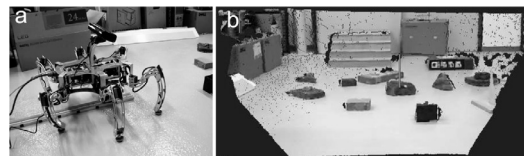
This paper addresses the problem of determining the egomotion between consecutive robot poses. The RGB-D Kinect sensor is used, which yields large amount of 3D points. It is impossible to process these data in real-time on most of the mobile robots. Thus, we present two approaches to point feature detection: 3D geometric features obtained from point clouds [9] and photometric 2D features detected in the RGB image [1]. Both methods are compared on publicly available RGB-D datasets [8, 10]. The detection on 2D image is the core of the currently state-of-the-art SLAM systems [3] while 3D features are an invention made especially for data captured by the Kinect-like sensors. First, it is demonstrated that both methods can be a part of a successful visual odometry system. Moreover, it is shown that detection of 2D image features is much faster than 3D, while the description is faster in case of 3D features (Tab. 1.). Performed experiments revealed that 3D geometric features tend to work better in environments of richer geometric structure (Tab. 2.) while method using photometric 2D features can be successfully expanded to additionally use 2D features detected on the depth image.

Keywords: mobile robots, Kinect sensor, feature detectors, self-localization.

1. Wstęp

Problem samolokalizacji robotów mobilnych jest problemem złożonym, rozważanym przez wielu badaczy. Jednym z istotnych kierunków badań są algorytmy odometrii wizyjnej, pozwalające na określenie przemieszczenia robota na podstawie informacji wizyjnej. Skuteczna metoda odometrii wizyjnej (ang. Visual Odometry - VO) może być także dobrą podstawą systemu jednoczesnej samolokalizacji i budowania mapy (ang. Simultaneous Localization and Mapping – SLAM) [4]. Odometria wizyjna może wykorzystywać różne sensory,

takie jak pojedyncza kamera, stereowizja, skaner laserowy 3D, czy też stosunkowo nowa klasa sensorów aktywnych RGB-D, takich jak Kinect. Wizja pasywna ma wiele praktycznych ograniczeń, a skanery laserowe 3D są stosunkowo ciężkie i duże. Dlatego aktywne sensory RGB-D są dobrym wyborem, gdy konieczne jest ograniczenie masy lub zużycia energii, co jest charakterystyczne np. dla robotów latających czy kroczących (rys. 1a). Dodatkowym atutem sensorów takich jak Kinect jest cena. Zasada działania sensora oparta jest na oświetleniu sceny z użyciem światła strukturalnego, czyli projekcji odpowiedniego wzoru geometrycznego (ang. structured light pattern) na obserwowane obiekty. Uzyskane ułożenie wzoru na scenie porównywane jest z obrazem referencyjnym zapisanym w pamięci urządzenia. Pozwala to na oszacowanie głębi sceny. Ze względu na konieczność kontrastowego oświetlenia sceny w zakresie bliskiej podczerwieni sensor nie może być skutecznie używany w warunkach ekspozycji na światło słoneczne mające silną składową podczerwoną. Sensor Kinect wyposażony jest również w standardową kamerę RGB o rozdzielczości 640x480 pixeli. Rozdzielczość obrazu głębi to 320x240 pixeli, jednak obraz ten jest interpolowany do rozmiaru obrazu z kamery, aby uzyskać odpowiedniość danych fotometrycznych (RGB) i geometrycznych (D) na poziomie poszczególnych pikseli. Pomiar głębi dla danego piksela opisany jest za pomocą 11-bitowej wartości, która dla skalibrowanego urządzenia określa mierzoną odległość w milimetrach. Wykorzystując obraz D z sensora Kinect, informację o parametrach wewnętrznych kamery oraz jej model otworkowy (ang. pin-hole) można wyliczyć reprezentację danych D w postaci chmury punktów w trzech wymiarach (rys. 1b). Maksymalna wielkość chmury to 307200 punktów.



rys. 1. Sensor Kinect na robocie Messor (a) oraz obraz RGB-D przykładowej sceny w laboratorium (b)

Fig. 1. Kinect sensor in walking robot Messor (a), and RGB-D data of an exemplary scene in the laboratory (b)

Możliwość uzyskania z sensora Kinect w czasie rzeczywistym „gęstej” mapy głębi sceny spowodowała powstanie nowych metod samolokalizacji robotów. Część badaczy skupiła się na przetwarzaniu pełnej chmury punktów i wykorzystaniu możliwości przetwarzania równoległego z użyciem procesorów graficznych (GPGPU). Takie rozwiązanie jest jednak niedostępne dla wielu robotów (np. kroczących) ze względu na ograniczenia sprzętowe.

Alternatywę stanowią metody oparte na przetwarzaniu wybranych, charakterystycznych punktów danego zbioru danych RGB-D, które następnie służą do wyliczenia transformacji między zbiorami danych uzyskanymi w różnych punktach. Efektem tego kierunku badań jest powstanie kilku systemów odometrii wizyjnej oraz SLAM. Reprezentatywny przykład rozwiązania należącego do tej grupy przedstawiono w [3].

Większość znanych metod punkty kluczowe znajduje na obrazach RGB i nie wykorzystuje dodatkowej informacji o strukturze mapy głębi sceny. Możliwość wykorzystania informacji o geometrycznej strukturze sceny dają detektory cech punktowych działające na chmurze punktów, takie jak NARF (ang. Normal Aligned Radial Features) [9]. W oparciu o punkty kluczowe wyznaczone ze zbioru danych 3D także można stworzyć system estymujący transformację łączącą pozycje robota [6].

W niniejszym artykule porównano własności metod detekcji cech punktowych w różnych podzbiorach zestawu danych RGB-D: na obrazie RGB, na obrazie D oraz bezpośrednio w chmurze punktów. Ocena własności została przeprowadzona ściśle w kontekście zastosowania metody ekstrakcji cech w systemie odometrii wizyjnej.

Praca została podzielona na trzy części: w pierwszej opisano zasadę działania metod estymujących transformację na podstawie położenia punktów kluczowych, i przedstawiono oryginalne modyfikacje omawianych metod, w drugiej eksperymentalnie zweryfikowano i porównano wybrane metody, a w trzeciej podsumowano wyniki i przedstawiono potencjalne, dalsze kierunki badań.

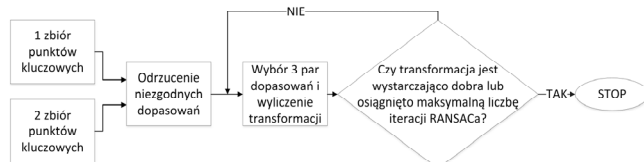
2. Zasada działania metod estymacji transformacji między pozycjami robota

Sensor RGB-D zamontowany na robocie mobilnym dostarcza informacji w postaci obrazu fotometrycznego RGB oraz obrazu głębi D. Dostęp do obrazu RGB pozwala na zastosowanie technik znanych z klasycznej odometrii wizyjnej [vo]. Do formalnego zapisu problemu poszukiwania transformacji między dwoma zbiorami punktów 3D oznaczone zostaną przez p_A oraz p_B odpowiednio punkty zbioru pierwszego oraz drugiego. Dodatkowo niech w_{ij} wynosi 1 w przypadku, gdy j -ty punkt zbioru p_B to i -ty punkt zbioru p_A widziany z innej pozycji robota. W przeciwnym wypadku punkt nie jest widoczny z obu rozważanych pozycji, a w_{ij} wynosi 0. Problem poszukiwania najlepszej transformacji można zapisać w postaci:

$$(\mathbf{R}, \mathbf{t}) = \operatorname{argmin}_{\mathbf{R}, \mathbf{t}} \left\{ \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} w_{i,j} \| \mathbf{p}_{A_i} - (\mathbf{R} * \mathbf{p}_{B_j} + \mathbf{t}) \|^2 \right\},$$

gdzie (\mathbf{R}, \mathbf{t}) są poszukiwanymi estymatami odpowiednio macierzy rotacji i wektora translacji, a n_A oraz n_B są liczbami określającymi licznosc zbiorów p_A oraz p_B .

W zaproponowanych metodach poszukiwanie optymalnej transformacji wykonuje się na podzbiorze wszystkich punktów kluczowych, określających położenie cech w przestrzeni. Potok przetwarzania wszystkich badanych metod zaczyna się od detekcji tych punktów (rys. 2).



Rys. 2. Schemat blokowy działania metody RANSAC w zadaniu estymacji transformacji

Fig. 2. Block diagram of RANSAC method used to estimate transformation

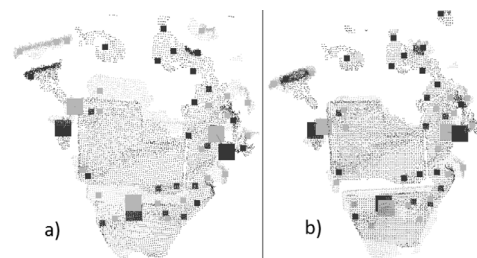
W niniejszej pracy skupiono się na zbadaniu etapu detekcji i deskrypcji punktów kluczowych w metodach odometrii wizyjnej RGB-D oraz określeniu wpływu rodzaju cech punktowych na jakość wyliczonych transformacji. W ramach pracy zbadano dwie różne metody, które wyszukiwały punkty kluczowe, o diametralnie różnej charakterystyce. W wersji algorytmu opartej na detekcji na obrazie RGB wykorzystano detektor SURF (ang. Speeded-Up Robust Features) [1], który charakteryzuje się dobrą skutecznością w stosunku do nakładów obliczeniowych. W wersji algorytmu wykorzystującej detekcję cech punktowych w chmurze punktów 3D wykorzystano dedykowany detektor NARF [9].

Punkty kluczowe NARF wykrywane są w chmurze punktów w przestrzeni, a zasada detekcji polega na dopasowywaniu płaszczyzny do potencjalnego punktu kluczowego oraz jego otoczenia. W przypadku, gdy dopasowana płaszczyzna posiada odpowiednie własności, badany punkt uznawany jest za punkt kluczowy. W przeciwieństwie do NARF, cechy SURF wykrywane są na obrazie 2D, zazwyczaj fotometrycznym (RGB), za pomocą badania histogramów lokalnych gradientów intensywności. Każdy punkt kluczowy można opisać za pomocą deskryptora, który jest

wektorem wartości liczbowych opisujących charakterystykę danego punktu kluczowego i jego otoczenia. Dla punktów NARF jest to opis charakterystyki płaszczyzny w pewnym otoczeniu punktu, a dla punktów SURF jest to opis zmian intensywności obrazu w ich otoczeniu. Z wykorzystaniem deskryptorów, opisane punkty kluczowe z kolejnych zbiorów danych RGB-D można porównywać i dopasowywać pod względem podobieństwa. Zbiór potencjalnie dopasowanych par punktów kluczowych może być użyty do znalezienia rzeczywistej transformacji. Niestety, nie wszystkie dopasowania punktów kluczowych są poprawne. W niniejszej pracy, w celu zwiększenia odporności na błędne dopasowania wykorzystano znaną metodę RANSAC [2]. Jest to sposób losowego przeglądania zbioru potencjalnych modeli i analizowania ich dopasowania do danych. W przypadku systemów samolokalizacji metoda RANSAC sprowadza się do losowego wyboru trzech par potencjalnych dopasowań, wyliczeniu modelu transformacji za pomocą algorytmu Kabscha [5] korzystającego z rozkładu SVD (ang. Singular Value Decomposition) oraz zbadaniu poprawności i jakości danej transformacji. Jakość transformacji mierzona jest stosunkiem dopasowań zgodnych z modelem do wszystkich dopasowań. Przez zgodne dopasowanie rozumiane jest dopasowanie, dla którego zastosowanie transformacji na punkcie ze zbioru p_A powoduje uzyskanie współrzędnych odpowiadającego mu punktu ze zbioru p_B z określoną dokładnością. W opisywanych badaniach założono, że dokładność ta wynosi 1 cm. Działanie metody RANSAC zostaje przerywane, gdy najlepsza transformacja (model) osiągnie jakość dopasowania, która jest powyżej pewnego progu, lub gdy osiągnięta zostanie maksymalna liczba iteracji. Zakładając określony procent poprawnych dopasowań oraz ustalając oczekiwaną skuteczność wyliczenia poprawnej transformacji można oszacować oczekiwaną liczbę iteracji RANSAC [2].

Ostatnim krokiem algorytmu jest reestymacja transformacji na podstawie wszystkich dopasowań, które zostały uznane za zgodne. Transformacja jest wyliczana, aby zminimalizować błąd średniokwadratowy dopasowań. W ten sposób uzyskuje się dodatkową odporność na szum pojedynczych punktów kluczowych. Ostateczna transformacja jest zapisywana jako wynik działania systemu odometrii wizyjnej. Tak znaną transformację można zastosować dla każdego punktu ze zbioru n_A .

Przykładowe zbiory danych, przed i po transformacji, przedstawiono we wspólnym układzie współrzędnych na rys. 3. Na Rys. 3a kolorem czerwonym oraz zielonym przedstawiono dwa zbiory danych, które mają zostać dopasowane. Małymi kwadratami oznaczono wykryte punkty kluczowe. Dużymi kwadratami oznaczono punkty kluczowe, które zostały wybrane za pomocą metodyki RANSAC do wyliczenia proponowanej transformacji. Rysunek 3b przedstawia efekt zastosowania wyznaczonej transformacji. Można zaobserwować nałożenie się zbiorów danych, co świadczy o znalezieniu transformacji, która jest bliska transformacji rzeczywistej.



Rys. 3. Wizualizacja chmur punktów: po wykryciu dopasowania punktów NARF (a) oraz po zastosowaniu wyliczonej transformacji (b)

Fig. 3. Point clouds visualization: after NARF keypoints matching (a), and after applying computed transformation (b)

3. Wyniki eksperymentalne

Wszystkie eksperymenty zostały przeprowadzone na laptopie wyposażonym w 2-rdzeniowy, 4-wątkowy procesor Intel Core i5-

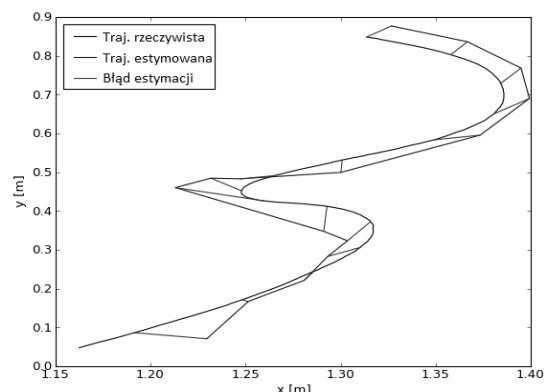
3210M oraz 4 GB pamięci RAM. Do żadnego z zadań opisanych w artykule nie wykorzystywano możliwości przetwarzania wielowątkowego czy przetwarzania równoległego na kartach graficznych (GPGPU). Stworzony program korzysta z otwartych bibliotek OpenCV w wersji 2.4.6 oraz PCL w wersji 1.7 wraz z bibliotekami zależnymi. Biblioteka OpenCV została wykorzystana do przetwarzania obrazów, natomiast biblioteka PCL służy do przetwarzania danych w postaci chmury punktów. Do eksperymentów wykorzystano dwa publicznie dostępne zbiory danych: zbiór pochodzący z Uniwersytetu we Freiburgu¹ [10] oraz zbiór danych PUT pochodzący z Politechniki Poznańskiej² [8].

Tab. 1. Porównanie czasów detekcji i deskrypcji punktów kluczowych 2D i 3D
Tab. 1. Comparison of detection and description times of 2D and 3D features

PUT1 (448 zdjęć)	Czas detekcji	Liczba punktów kluczowych	Czas deskrypcji
SURF (hessian =10)	$\mu = 0,046320$ s $\sigma = 0,0077543$ s	$\mu = 1584,5$ $\sigma = 322,43$	$\mu = 0,11299$ s $\sigma = 0,027452$ s
SURF (hessian =50)	$\mu = 0,035219$ s $\sigma = 0,0055842$ s	$\mu = 1006,6$ $\sigma = 310,84$	$\mu = 0,081304$ s $\sigma = 0,025738$ s
SURF (hessian =100)	$\mu = 0,034265$ s $\sigma = 0,0064774$ s	$\mu = 758,73$ $\sigma = 286,52$	$\mu = 0,066822$ s $\sigma = 0,024222$ s
SURF (hessian =400)	$\mu = 0,030203$ s $\sigma = 0,0046417$ s	$\mu = 413,17$ $\sigma = 222,72$	$\mu = 0,043888$ s $\sigma = 0,021647$ s
NARF (roz. kąt = 0.1)	$\mu = 2,1321$ s $\sigma = 0,11282$ s	$\mu = 993,62$ $\sigma = 585,80$	$\mu = 0,054229$ s $\sigma = 0,032365$ s
NARF (roz. kąt = 0.2)	$\mu = 0,62143$ s $\sigma = 0,030860$ s	$\mu = 252,20$ $\sigma = 107,38$	$\mu = 0,013229$ s $\sigma = 0,0056688$ s
NARF (roz. kąt = 0.5)	$\mu = 0,19604$ s $\sigma = 0,0067492$ s	$\mu = 75,482$ $\sigma = 28,805$	$\mu = 0,0032202$ s $\sigma = 0,0011923$ s
NARF (roz. kąt = 1)	$\mu = 0,14772$ s $\sigma = 0,010852$ s	$\mu = 40,004$ $\sigma = 11,788$	$\mu = 0,0017403$ s $\sigma = 0,0005232$ s

W tabeli 1 zebrano porównanie czasów detekcji i deskrypcji obu metod dla danych z bazy PUT. Do testów wykorzystano sekwencję PUT1, która zawiera 448 obrazy. Czas detekcji uśredniono na podstawie wszystkich zdjęć należących do danej sekwencji. Dla każdej sekwencji wykonano 4-krotną detekcję detektorem NARF oraz SURF z różnymi parametrami detekcji, sterującymi liczbą uzyskanych punktów kluczowych. Dla detektora NARF parametrem sterowanym był promień sfery traktowanej jako otoczenie punktu oraz rozdzielczość kątowna detektora. Badano parametry otoczenia wynoszące odpowiednio: 0,2; 0,4 oraz 0,8, jednak nie zaobserwowano zmiany czasu detekcji, ani zmiany liczby wykrytych punktów kluczowych. Wyniki dla rozdzielczości kątownej (w tabeli określonej skrótowo roz. kąt) wynoszącej 0,1; 0,2, 0,5 oraz 1 stopień przedstawiono w tabeli 1. Dla detektora SURF badano wpływ progu wartości hessianu (w tabeli oznaczonej jako "hessian") dla wartości: 10, 50, 100 oraz 400. Przeprowadzone badania wykazały, że dla detektora SURF liczba znalezionych cech ma nieznaczny wpływ na czas detekcji. Przeciwny efekt można zaobserwować dla detektora NARF, gdy dla większej liczby punktów kluczowych otrzymuje się większy czas obliczeń. Dla zbioru danych PUT1 okazało się, że metoda oparta na detektorze SURF jest zdecydowanie szybsza od metody opartej na punktach kluczowych NARF. Dodatkowo można zauważyć, że metoda SURF dla podobnego czasu detekcji znajduje więcej punktów kluczowych niż metoda NARF.

Dodatkowo zbadano także czasy deskrypcji w zależności od liczby wykrytych punktów kluczowych. Każdy pomiar czasu deskrypcji przeprowadzono na 448 zdjęciach, dla których uśredniono uzyskane czasy. Badania wykazały, że czas deskrypcji dla obu metod jest w przybliżeniu liniowo zależny od liczby punktów kluczowych, które zostają opisane. Zakładając prawdziwość liniowej zależności, można oszacować, że czas deskrypcji pojedynczego punktu dla metody NARF wynosi około 0,05 ms, a dla metody SURF jest to około 0,1 ms.



Rys. 4. Trajektorja uzyskana metodą SURF dla fragmentu ze zbioru freiburg_room
Fig. 4. Trajectory estimated by the SURF method for the part of the freiburg_room dataset

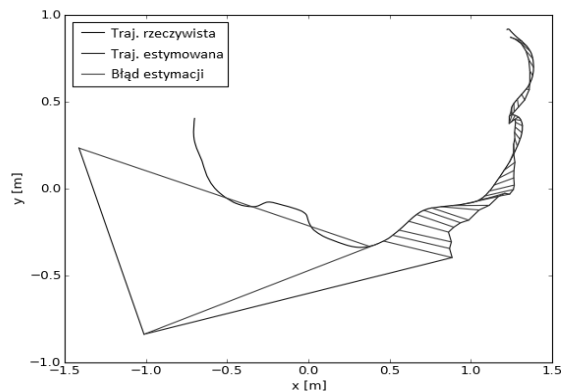
Następnym przeprowadzonym testem była próba pomiaru dokładności rekonstrukcji transformacji. Estymowana trajektoria ruchu robota powstawała poprzez złożenie transformacji z dopasowań kolejnych par klatek obrazu. Testy przeprowadzono dla 2 trajektorii ze zbioru danych pochodzącego z PUT oraz dla 2 trajektorii pochodzących z Uniwersytetu we Freiburgu. Analizę dokładności przeprowadzono korzystając z narzędzi udostępnionych wraz ze zbiorem danych z Freiburga [10]. Dokładność zmierzono za pomocą narzędzia ATE (ang. Absolute Trajectory Error) mierzącego absolutny błąd trajektorii.

Na rys. 4. przedstawiono trajektorię powstałą przez dopasowanie 11 par danych pochodzących z sekwencji freiburg_desk1 dzięki zastosowaniu metody opartej na detekcji SURF. Korzystając z narzędzia ATE zmierzono także średniokwadratowy błąd trajektorii, który wyniósł 0,040526 m. Dla danej sekwencji, minimalny błąd wyniósł 0,009194 m, natomiast maksymalny 0,067753 m. Następnym krokiem była próba użycia metody NARF na tym samym fragmencie sceny w celu porównania obu metod. Okazało się, że dla metody NARF praktycznie niemożliwe jest dobranie nastaw, które pozwolą na poprawne wykrywanie transformacji dla dłuższych sekwencji danych. Najdłuższą sekwencją, dla której sprawdziły się jedne nastawy metody NARF była sekwencja 4 par danych. Dla takiej sekwencji błąd średniokwadratowy wyniósł 0,023734 m. Błąd minimalny wyniósł 0,016429 m, a maksymalny 0,033561 m. Niestety, żeby otrzymać takie nastawy konieczne było wykrycie około 400-500 punktów kluczowych, co dla jednej chmury punktów zajmuje około 1,2 s. Długi czas działania oraz trudność w doborze odpowiednich parametrów detekcji powodują, że metoda ma mały potencjał, jeśli chodzi o potencjalną aplikowalność w systemie SLAM.

W ramach badań podjęto także próbę zastosowania metod dla dłuższych sekwencji. Zgodnie z oczekiwaniami dobór odpowiednich parametrów dla metody NARF okazał się zadaniem niemożliwym do realizacji. Większa liczba wykrytych punktów kluczowych spowodowała problem z wykrywaniem dopasowań pomiędzy punktami kluczowymi zbiorów danych. Ujawnia to słabość deskryptora NARF, który nie jest wystarczająco unikatowy. Jedyną możliwością kompensacji małego udziału poprawnych dopasowań dla badanych modeli transformacji jest znaczne zwiększenie liczby iteracji RANSAC, ale jednocześnie powoduje to zwielokrotnienie nakładów obliczeniowych i brak możliwości użycia na robocie mobilnym. Na rys. 5 przedstawiono rekonstrukcję dłuższej sekwencji dla metody opartej na cechach SURF. Widać, że pomimo małych błędów dopasowania poszczególnych par danych, sumaryczny błąd dopasowania jest kumulowany i błąd bezwzględnej trajektorii rośnie bardzo szybko. Dlatego proste dopasowywanie kolejnej informacji do poprzedniej pozwala na rekonstrukcję trajektorii jedynie dla krótkiego okna czasowego np. 9-16 dopasowań. W sytuacji, gdy zależny nam na uzyskaniu dłuższych trajektorii konieczne staje się zastosowanie technik optymalizacji znanych z systemów SLAM [3], które pozwalają na uzyskanie globalnie spójnej mapy otoczenia.

¹ <http://vision.in.tum.de/data/datasets/rgbd-dataset>

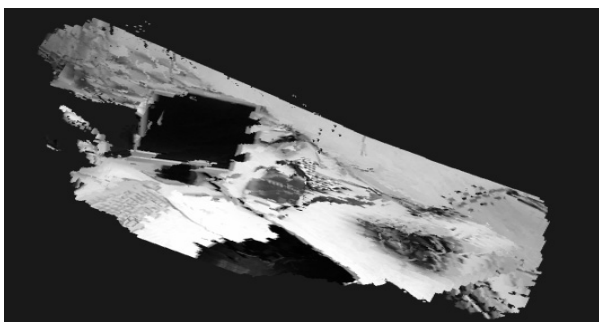
² <http://www.vision.put.poznan.pl/>



Rys. 5. Trajektorie metody opartej na detektorze SURF, dla której obserwuje się dryft estymaty oraz niepoprawną estymację dla fragmentu ubogiego w cechy fotometryczne

Fig. 5. Trajectories estimated by the method based on the SURF detector with significant estimation drift and wrong estimation for the scene lacking in the photometric features

Kolejny eksperyment przeprowadzono w środowisku o małej zawartości cech fotometrycznych. Taka sytuacja miała miejsce na fragmencie sceny z zbioru danych PUT1, gdy scena zawierała białe obiekty na białym tle, a trajektoria charakteryzowała się dużym obrotem wynoszącym $15,39^\circ$ w osi Z robota. Detektor NARF dla opisanej sceny uzyskała zadowalający wynik ($12,39^\circ$) będący dobrą aproksymacją rzeczywistej transformacji. Dla takiego zbioru danych detekcja punktów kluczowych SURF nie pozwoliła na znalezienie dobrych punktów kluczowych i ostatecznie na oszacowanie transformacji bliżej rzeczywistej ($0,32^\circ$). Wynika to z faktu, że metoda korzystająca z detekcji i deskrypcji SURF na obrazach RGB nie wykorzystuje w pełni wiedzy o obrazie głębi pochodzącym z sensora Kinect. Dlatego postanowiono wzbogacić tę metodę o detekcję i deskrypcję na obrazach głębi. Uzyskano w ten sposób dwa zbiory potencjalnych dopasowań. Następnie tworzono możliwe dopasowania w ten sam sposób jak poprzednio, jednak przy założeniu, że punkty kluczowe z obrazu RGB mogą być dopasowane jedynie do punktów kluczowych z obrazu RGB. Analogicznie postąpiono dla obrazu głębi. Pomimo faktu, że cechy geometryczne mają inną charakterystykę niż cechy fotometryczne, do detekcji na obrazie głębi postanowiono skorzystać z detektora FAST (ang. Features from Accelerated Segment Test) [7], który pierwotnie został zaproponowany do detekcji na obrazie RGB. Tak stworzona metoda ma w stosunku do standardowej metody SURF zwiększoną liczbę potencjalnych dopasowań i lepiej sprawdziła się dla środowiska o bogatej strukturze geometrycznej ($4,52^\circ$).



Rys. 6. Wizualizacja 15 kolejnych chmur punktów ze zbioru freiburg_desk zgodnie z transformacjami wyliczonymi przez metodę opartą na punktach SURF

Fig. 6. Visualization of 15 consecutive point clouds from freiburg_desk dataset transformed accordingly to the transformations estimated by method using SURF keypoints

Wyliczone transformacje można także wykorzystać do wizualizacji kolejnych przetwarzanych zbiorów danych. Na podstawie estymowanych transformacji można dokonać transformacji zbiorów do wspólnego układu współrzędnych,

a następnie wspólnie przedstawić je na jednym rysunku. Taką sytuację przedstawiono na rys. 6, gdzie do wykonania rysunku wykorzystano 15 kolejnych zbiorów danych z sekwencji freiburg_desk. Do wyliczenia estymat transformacji wykorzystano metodę opartą na detekcji SURF. Warto tutaj zaznaczyć, że otrzymany rysunek powstał bez dodatkowego usuwania nieciągłości obiektów czy optymalizacji chmur punktów.

4. Wnioski

Przeprowadzone testy wykazały, że detekcja cech na obrazach 2D jest metodą szybszą i skuteczniejszą w zadaniu samolokalizacji niż detekcja cech na danych 3D. Poprawna implementacja metod 2D powinna pozwolić na uzyskanie możliwości pracy w czasie rzeczywistym nawet na robocie o ograniczonych możliwościach przetwarzania. System oparty na detektorze NARF jest zbyt wolny, aby przy liczbie punktów kluczowych wystarczających do skutecznej samolokalizacji móc działać w czasie rzeczywistym oraz zbyt wrażliwy na zmiany charakteru środowiska aby dobrać stałe parametry dla dłuższych trajektorii ruchu robota. Jednak badania przeprowadzone ze sceną zawierającą niewiele cech fotometrycznych sugerują, że w przypadku danych z sensora Kinect warto wykorzystać dodatkową wiedzę o strukturze sceny. Prostim i skutecznym pomysłem jest wykorzystanie znanych detektorów punktowych cech fotometrycznych do wykrycia cech także na obrazie głębi. Takie rozwiązanie pozwala na uzyskanie większej odporności na potencjalne "zgubienie się" metody samolokalizacji w sytuacji braku wystarczającej liczby cech na obrazie RGB. Następnym krokiem prac będzie próba odpowiedzi na pytanie, które ze znanych detektorów są najlepszym wyborem w przypadku detekcji punktów kluczowych na obrazie głębi. W przypadku uzyskania niezadowolających wyników, alternatywnym krokiem prac jest stworzenie własnego, szybkiego detektora pracującego na obrazie głębi.

5. Literatura

- [1] Bay H., Ess A., Tuytelaars T., Van Gool L.: Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.*, vol. 110(3), p. 346-359, 2008.
- [2] Choi S., Kim T., Yu W.: Performance evaluation of RANSAC family. *Proc. British Machine Vision Conference*, p. 81.1-81.12, 2009.
- [3] Endres F., Hess J., Engelhard N., Sturm J., Cremers D., Burgard W.: An Evaluation of the RGB-D SLAM System. *Proc. IEEE Int. Conf. on Robot. and Automat.*, p. 1691-1696, 2012.
- [4] Fraundorfer F., Scaramuzza D.: Visual odometry: Part II - matching, robustness and applications. *IEEE Robotics & Automation Magazine*, 19(2), p. 78-90, 2012.
- [5] Kabsch W.: A solution of the best rotation to relate two sets of vectors. *Acta Crystallographica* 32:922, 1976.
- [6] Nowicki M., Skrzypczyński P.: Eksperymentalna weryfikacja metody samolokalizacji robota kroczonego z sensorem Kinect. *Prace Naukowe Politechniki Warszawskiej*, vol. 2, p. 561-572, 2012.
- [7] Rosten E., Drummond T.: Machine learning for high-speed corner detection. *Proc. European Conference on Computer Vision*, vol. 1, p. 430-443, 2006.
- [8] Schmidt A., Fularz M., Kraft M., Kasiński A., Nowicki M.: An indoor RGB-D dataset for the evaluation of robot navigation algorithms. *Lecture Notes in Computer Science*, vol. 8192, p. 321-329, 2013.
- [9] Steder B., Rusu R. B., Konolige K., Burgard W.: Point feature extraction on 3D range scans taking into account object boundaries. *Proc. IEEE Int. Conf. on Robot. and Automat.*, p. 2601-2608, 2011.
- [10] Sturm J., Engelhard N., Endres F., Burgard W., Cremers D.: A Benchmark for the Evaluation of RGB-D SLAM System. *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, p. 573-580, 2012.