Akwizycja obrazów RGB-D: metody

Maciej Stefańczyk, Tomasz Kornuta

Instytut Automatyki i Informatyki Stosowanej, Politechnika Warszawska

Streszczenie: Dwuczęściowy artykuł poświęcono czujnikom umożliwiającym akwizycję chmur punktów oraz map głębi. W poniższej, pierwszej części uwagę skupiono na trzech głównych metodach pomiarowych: stereowizji, świetle strukturalnym oraz pomiarze czasu lotu wiązki jako tych, które są najpowszechniej stosowane w robotyce. Poza zasadą działania każdej z metod przeanalizowano także ich właściwości, złożoność obliczeniową oraz potencjalne zastosowania.

Słowa kluczowe: obraz RGB-D, czujnik RGB-D, mapa głębi, chmura punktów, czas lotu wiązki, światło strukturalne, stereowizja

DOI: 10.14313/PAR_203/82

1. Wprowadzenie

Otrzymanie informacji opisującej scenę oraz obiekty się na niej znajdujące jest naczelnym celem wizji komputerowej od chwili powstania tej gałęzi nauki. Kombinacja informacji kolorowej z mapą głębi z jednej strony umożliwia przezwyciężenie szeregu klasycznych problemów wizji komputerowej, ale z drugiej tworzy nowe problemy i wyzwania. Zainteresowanie tą tematyką objawia się m.in. podczas licznych warsztatów i sesji specjalnych poświęconych stosowaniu czujników RGB-D na największych międzynarodowych konferencjach dotyczących robotyki oraz wizji komputerowej. Przykładami są warsztaty *RGB-D: Advanced Reasoning with Depth Cameras* organizowane rokrocznie przy konferencji Robotics: Science and Systems (RSS) [7–9, 22], sesja specjalna *3D Point Cloud Processing: PCL* na konferencji IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) [20], *RGB-D Workshop on 3D Perception in Robotics* na Europejskim Forum Robotycznym (ERF) [1], warsztaty *IEEE Workshop on Consumer Depth Cameras for Computer Vision* organizowane przy konferencjach ICCV/ECCV [4–6], czy też sesja specjalna *Percepcja robotów z wykorzystaniem obrazów RGB-D*, która zostanie zorganizowana na 13. Krajowej Konferencji Robotyki (KKR) [15]. Warto również zwrócić uwagę na poświęcone tej tematyce numery specjalne renomowanych międzynarodowych czasopism [10, 23].

Postęp na tym polu nie byłby możliwy bez rozwoju czujników umożliwiających akwizycję obrazów kolorowych wraz z mapami głębi. Celem dwuczęściowego artykułu jest przegląd aktualnie dostępnych na rynku rozwiązań. W pierwszej części skupiono uwagę na metodach pozyskiwania obrazów RGB-D. Pomimo istnienia szeregu technik, za pomocą których można otrzymać głębię, np. przez analizę strukturalną cienia (ang. *Depth From Shading*) lub analizę ruchu (ang. *Depth From Motion*), uwagę skupiono na trzech z nich: stereowizji, świetle strukturalnym oraz pomiarze czasu lotu wiązki. Wybór ten wynika głównie z dominacji tego typu czujników w aplikacjach robotycznych, a także ich dostępności na rynku. W drugiej części artykułu omówiono obecne na rynku komercyjne, sprzętowe rozwiązania wraz z ich krótką charakterystyką.

2. Nomenklatura

Istnieje szereg technik umożliwiających akwizycję informacji przestrzennej z obserwowanej sceny. Techniki te, a co za tym idzie również i czujniki, można podzielić na dwie główne klasy: aktywne i pasywne. Działanie czujni-

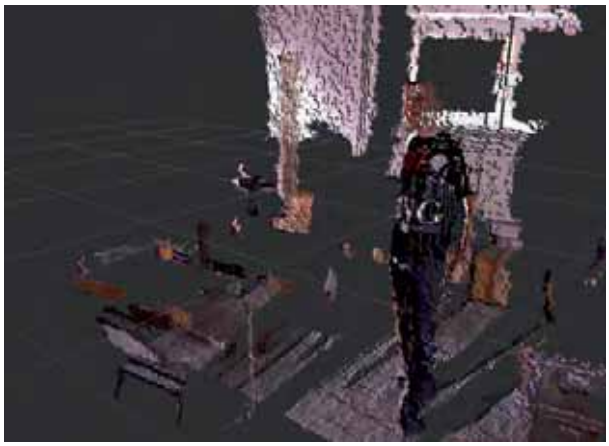
Rys. 1. Mapa głębi uzyskana bezpośrednio z urządzenia obrazującego (Kinect); jasność oznacza odległość punktu od kamery – im jaśniejszy, tym punkt znajduje sie dalej, punkty całkowicie czarne oznaczają brak odczytu

Fig. 1. An exemplary depth map gathered from the imaging device (Kinect). The brighter the point is the farther from the camera it is. Black colour indicates the lack of proper distance measure

ków aktywnych związane jest z emisją dodatkowej energii do środowiska (np. oświetlenie dodatkowym światłem). Czujniki pasywne z kolei bazują jedynie na pasywnie otrzymanej od środowiska energii (np. fotonach, które dotrą do ich sensora).

Główną pasywną techniką otrzymywania informacji przestrzennej jest stereowizja, której działanie stanowi analogie ludzkiego systemu percepcji wizyjnej. Celem tej techniki jest jednoczesne pozyskanie dwóch obrazów z pary kamer, a następnie obliczenie tzw. niezgodności (przesunięć w obrazach) między odpowiadającymi sobie pikselami. Informacja ta zapisywana jest w postaci tzw. mapy niezgodności (ang. *Disparity Map*) i w konsekwencji umożliwia otrzymanie mapy głębi (ang. *Depth Map*, rys. 1). Mapa taka może być i bardzo często jest traktowana jak monochromatyczny (jednokanałowy) obraz, którego każdy punkt przechowuje odległość do obiektu w kierunku wyznaczonym przez półprostą wychodzącą z optycznego środka urządzenia obrazującego i przechodzącą przez konkretny piksel na matrycy.

Można wyróżnić dwa rodzaje map głębi: mapy gęste (ang. *Dense Depth Map*), w których (prawie) każdy piksel obrazu przechowuje informację o głębi, oraz rzadkie (ang. *Sparse Depth Map*), w których tylko niektóre mają taką informację. Związane to głównie jest z metodą ekstrakcji mapy głębi – gęste mapy otrzymywane są np. przez analizę całego obrazu pochodzącego z kamery, natomiast rzadkie przez analizę jedynie pewnych punktów charakterystycznych (np. wierzchołków czy krawędzi).



Rys. 2. Trójwymiarowa wizualizacja obrazu RGB-D. Widok z innego punktu niż punkt widzenia urządzenia obrazującego (Kinect)

Fig. 2. A three-dimensional visualization of an RGB-D image. The presented point of view differs from the pose of the imaging device (Kinect)

Obrazem RGB-D nazywamy natomiast strukturę danych złożona z obrazu kolorowego wraz ze skojarzona mapa głębi (rys. 2). Jest to więc obraz posiadający cztery kanały: trzy związane z natężeniem koloru czerwonego, zielonego i niebieskiego oraz czwarty kanał związany z głębią. Możliwe jest także pozyskanie obrazów, gdzie zamiast pełnej informacji o kolorze (RGB) przechowywana jest jedynie intensywność (I). Obrazy te będą nazywane obrazami I-D. Warto dodać, iż możliwe jest dalsze zwiększanie ilości informacji przechowywanej w obrazie, np. przez dodawanie kanałów przechowujących wektory normalne czy współczynniki krzywizny powierzchni w danym punkcie [24], co może być użyteczne w procesie rozpoznawania.

Rys. 3. Chmura punktów zintegrowana z kilkudziesięciu odczytów laserowych [www.igi.eu]

Fig. 3. An exemplary point cloud integrated from multiple laser scans [www.igi.eu]

Alternatywnie, informacja o głębi może również być przechowywana w tzw. chmurze punktów (ang. *Point Cloud*). Przykładową chmurę punktów pokazano na rys. 3. W reprezentacji tej każdy punkt jest w istocie punktem w przestrzeni kartezjańskiej (ma swoje współrzędne), opisanym często za pomocą dodatkowych danych (kolor, współrzędne wektora normalnego do powierzchni itp.), co ułatwia m.in. fuzję chmur punktów z kolejnych chwil czasowych lub z różnych czujników. Zastosowanie chmur punktów umożliwiło także zupełnie inne spojrzenie na informację przestrzenną, czego konsekwencją jest rozwój szeregu nowych technik analizy i rozpoznawania obiektów oraz scen, a także adaptacja tradycyjnych algorytmów analizy obrazów 2D do 3D.

Obie reprezentacje głębi są częściowo kompatybilne oraz istnieją sposoby przetwarzania jednej reprezentacji w drugą. W szczególności, z każdej mapy głębi można uzyskać chmurę punktów, w druga stronę – nie zawsze jest to wykonywane bezstratnie. Wiąże się z tym jeszcze jedno pojęcie – tzw. uporządkowanie. Uporządkowaną chmurą punktów (ang. *Ordered Point Cloud*) nazywana jest chmura przechowująca punkty w postaci dwuwymiarowej tablicy, która powstaje zazwyczaj z przekształcenia mapy głębi. W tej reprezentacji punkty położone blisko siebie, w tablicy są położone również blisko w przestrzeni kartezjańskiej, natomiast punkty oddalone od siebie w tablicy, znajdują się daleko od siebie w rzeczywistości.

Po połączeniu dwóch lub większej liczby chmur punktów, zwykle nie jest możliwe utrzymanie ich uporządko-

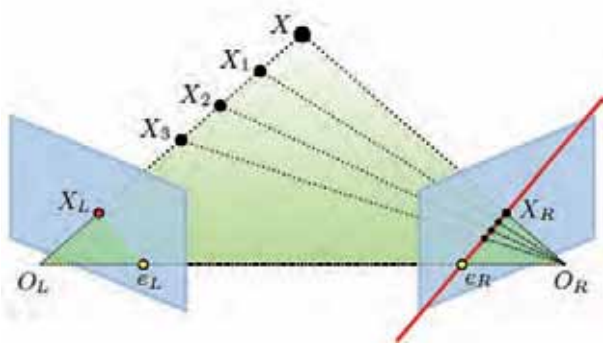
wania, wówczas mówi się o chmurze nieuporządkowanej (ang. *Unordered Point Cloud*). Ponieważ w takiej chmurze utrudnione jest przeszukiwanie sąsiedztwa, dlatego często do jej przechowywania stosowane są inne struktury danych, np. drzewa ósemkowe. Jest to jednak reprezentacja zupełnie niezależna od metody pozyskiwania informacji trójwymiarowej.

3. Stereowizja

3.1. Zasada działania

Stereowizja (ang. *Stereovision*) jest techniką obrazowania opierającą się na analizie obrazów pochodzących z wielu (najczęściej z dwóch) kamer. Obliczenie głębi bazuje na dysparycji, czyli względnej odległości między obrazami tego samego punktu w różnych kamerach. Można wydzielić trzy główne etapy stereowizji:

1) detekcja punktów charakterystycznych,
2) dopasowanie odpowiedników,
3) rekonstrukcja współrzędnych 3D.

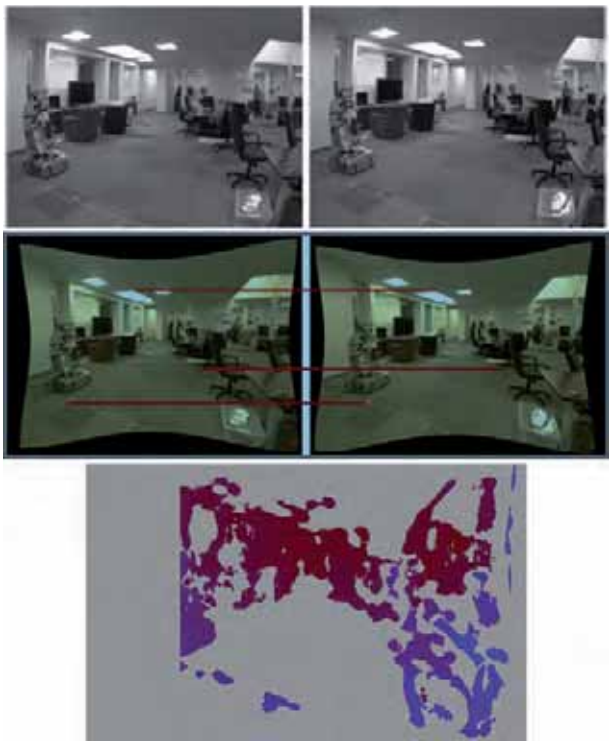


Rys. 4. Geometria dwubiegunowa [en.wikipedia.org/wiki/Epipolar_geometry]

Fig. 4. The principle of epipolar geometry [en.wikipedia.org/wiki/Epipolar_geometry]

Na rys. 4 przedstawiono poglądowo zasadę działania algorytmów stereowizyjnych. Rzeczywiste punkty X, X_1, X_2 oraz X_3 są współliniowe względem lewej kamery (widoczne w obrazie jako ten sam punkt X_L). W obrazie z prawej kamery punkty te są juz rozróżnialne (punkt X jest widoczny jako punkt X_R). Znając współrzędne środków optycznych (O_L, O_R) i orientację obu kamer, można wyznaczyć współrzędne linii epipolarnych dla badanych punktów, a także rzeczywiste współrzędne tych punktów (na podstawie ich współrzędnych w obu obrazach). W stereowizji kamery najczęściej ustawia się tak, aby ich osie optyczne były równoległe, dzięki czemu można łatwo wyznaczyć linie epipolarne (które w takim przypadku będą poziome) oraz punkty charakterystyczne.

Dodatkowo konieczne jest wstępne przetworzenie obrazów, by przedstawiały widok w taki sposób, jakby płaszczyzny obrazowania kamer były równoległe. Proces ten zwany jest rektyfikacją obrazu (ang. *Image Rectification*). Wymaga on wstępnej kalibracji układu kamer, w wyniku której wyznaczana jest pozycja kamer względem siebie (przesunięcie i obrót) oraz parametry wewnętrzne każdej z kamer (długości ogniskowych i parametry związane ze zniekształceniami wnoszonymi m.in. przez soczewki obiektywu). Proces kalibracji przeprowadza się raz dla danego położenia kamer, po każdorazowej zmianie ich pozycji konieczna jest powtórna kalibracja. Na rys. 5 przedstawiono kolejne kroki ekstrakcji głębi na podstawie obrazów otrzymanych z dwóch kamer.



Rys. 5. Kolejne etapy stereowizji; 1: obraz pobrany bezpośrednio z kamer; 2: obraz po interpolacji kolorów i rektyfikacji, zaznaczono przykładowe dopasowane punkty charakterystyczne; 3: wynikowa mapa głębi dla danej sceny [wiki.ros.org/stereo_image_proc]

Fig. 5. Intermediate stereovision steps. 1: images gathered from cameras; 2: rectified images with interpolated colors, with exemplary feature points matched; 3: resulting depth map [wiki.ros.org/stereo_image_proc]

3.2. Wymagania sprzętowe

Najważniejszym elementem w stereowizji są kamery – otrzymane wyniki zależą bezpośrednio od ich jakości. Stosowanie kamer z interfejsem analogowym jest możliwe, jednak nie jest zalecane przy dynamicznych scenach. Obiekty poruszające się z dużą szybkością są wykrywane błędnie lub całkowicie ignorowane ze względu na występujące rozmycie wynikające ze stosowania przeplotu w procesie akwizycji. Najlepsze wyniki uzyskuje się, stosując kamery dobrej jakości z interfejsami cyfrowymi, które są pozba-

wione tej wady, jednak ich koszt jest dużo wyższy. Rozdzielczość uzyskiwanej głębi zależy bezpośrednio od rozdzielczości kamer, jednak wraz z jej wzrostem rośnie czas wymagany do przetworzenia pojedynczej klatki obrazu.

Istotny wpływ ma też rozstaw kamer (ang. *Baseline*) – kamery umieszczone blisko siebie będą dawały dobrą aproksymację głębi dla obiektów znajdujących się blisko nich, natomiast kamery rozmieszczone szerzej pozwalają na uzyskanie lepszej rozdzielczości głębi dla obiektów znajdujących się daleko, kosztem częściowej lub całkowitej utraty informacji o obiektach bliskich.

Problemem może być samo mechaniczne mocowanie kamer. Musi być wykonane bardzo solidnie, gdyż w przypadku nawet małej zmiany orientacji kamer względem siebie, wymagana jest ponowna kalibracja całego systemu. Aby pozbyć się tej wady można stosować zintegrowane moduły zawierające dwie (lub więcej) kamery w jednej obudowie, to jednak uniemożliwia eksperymentowanie z odległością kamer od siebie. Najpopularniejsze obecnie kamery do stereowizji zostaną przedstawione w drugiej części artykułu.

3.3. Złożoność obliczeniowa

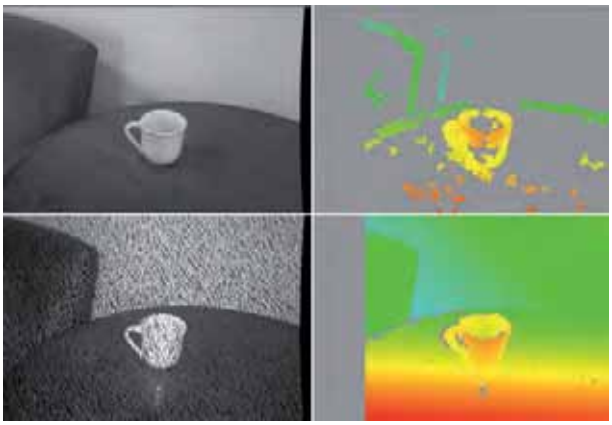
Stosując programową wersję algorytmów stereowizyjnych na standardowym komputerze domowym, można osiągnąć wydajność od kilku do kilkunastu klatek na sekundę [12, 26]. W związku z tym rozwiązania programowe nie nadają się do wykorzystania w środowisku, które zmienia się często i dynamicznie (takie zwykle jest otoczenie, w którym pracują roboty).

Dużo lepiej sprawdzają sie rozwiązania sprzętowe, w których algorytm tworzenia mapy niezgodności zaimplementowany jest w układach FPGA zintegrowanych w jednym module z kamerami. W tym przypadku wydajność jest stała i niezależna od platformy, na której uruchomione będą algorytmy sterowania robota, i wynosi (w zależności od producenta) od kilkunastu do ponad 30 FPS. Największą wadą takiego rozwiązania jest jego koszt – wynoszący od kilkuset do kilku tysięcy dolarów. Dla porównania, dwie kamery analogowe można kupić za ok. 200 $.

3.4. Zastosowania

Stereowizja bazuje często na analizie obrazu krawędziowego. W związku z tym obiekty o jednolitej, drobnej teksturze (lub całkowicie gładkie) są słabo lub całkowicie niewykrywalne. W najlepszym wypadku wykrywane są jedynie ich krawędzie, co prowadzi do powstawania dużych, niezidentyfikowanych obszarów w obrazie. Jednym z rozwiązań tego problemu może być zastosowanie dodatkowego projektora wyświetlającego specjalnie przygotowany wzór (rys. 6) pokrywający obiekty sztuczną teksturą umożliwiającą poprawę wyników stereowizji. Drugim sposobem poprawy sytuacji jest stosowanie dodatkowego etapu przetwarzania obrazu po wygenerowaniu wstępnej mapy głębi. Po segmentacji obrazu na podstawie koloru wybierane są obszary jednolite, których głębia interpolowana jest na podstawie głębi ich krawędzi.

Obydwa rozwiązania dają dobre efekty, jednak komplikują budowę urządzenia lub wprowadzają dodatkowy narzut obliczeniowy. Projekcja tekstury jest z powodzeniem stosowana np. na robotach manipulacyjno-przemysłowych, które są wyposażone w odpowiednio wydajne



Rys. 6. Przykład projekcji tekstury dla poprawy jakości stereowizji. W kolejności z góry od lewej: scena bez dodatkowego oświetlenia, wygenerowana mapa głębi, scena z rzutowaną teksturą, poprawiona mapa głębi [13]

Fig. 6. An example of the projected texture stereovision. From top left: scene without projected texture, its sparse depth map, scene with additional texture projected, resulting dense depth map [13]

jednostki obliczeniowe (np. robot PR2 [13]). Natomiast w robotach poruszających się w naturalnym środowisku, gdzie występuje bardzo dużo szczegółów (a więc i punktów charakterystycznych), stereowizja nie wymaga praktycznie żadnych dodatkowych usprawnień i sprawdza się bardzo dobrze (zwracane mapy głębi są wypełnione w ponad 80 %, a algorytm działa z szybkością powyżej 10 FPS) [14].

Istotną zaletą układów stereowizyjnych w zastosowaniu do pozyskiwania obrazów RGB-D jest idealne wyrównanie mapy głębi z obrazem kolorowym. Mapa niezgodności otrzymywana jest w układzie jednej z kamer, a obliczona na jej podstawie głębia pokrywa sie dokładnie z krawędziami obiektów. Nie ma też efektu tzw. cienia (braku głębi wokół obiektu), występującego w czujnikach pracujących w świetle strukturalnym. Z drugiej strony, ponieważ tylko część obrazu w obu kamerach jest wspólna, tylko dla niej można wyznaczyć głębię. Powoduje to zmniejszenie efektywnego pola widzenia.

4. Światło strukturalne

4.1. Zasada działania

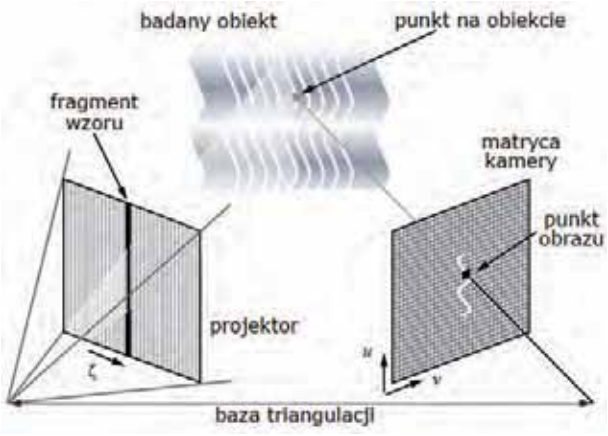
Inną metodą pomiaru i odtwarzania informacji o głębi sceny bazującą na analizie obrazu jest wykorzystanie światła strukturalnego (ang. *Structured Light*). Na scenę rzucane jest światło formujące znany wzorzec (ang. *Pattern*),

a kamera umieszczona jest w taki sposób, aby obserwować scenę pod innym kątem niż orientacja rzutnika. Na podstawie odczytanej deformacji wzorca za pomocą algorytmów bazujących na triangulacji wyliczane są rzeczywiste współrzędne punktów w obrazie.

Rzutowane mogą być różne wzorce, zaczynając od pojedynczego punktu, przez wzorce złożone z linii (statycznych lub przesuwających się po scenie), aż po złożone, pseudolosowe wzorce (monochromatyczne lub kolorowe) oraz sekwencje wzorców [19]. W przypadku korzystania z sekwencji wzorców wymagany jest statyczny charakter sceny – sceny dynamiczne wymagają stosowania pojedynczych, skomplikowanych wzorców [21].

4.2. Zastosowania

Głównym kryterium przy wyborze konkretnej realizacji skanera opartego na świetle strukturalnym jest charakter analizowanej sceny. W przypadku skanowania obiektów statycznych (np. podczas automatycznego tworzenia modeli trójwymiarowych) możliwe jest zastosowanie sekwencji wzorców, np. prążków Graya (rys. 8) lub prążków De Bruijna [11]. Stosowane są też różne rozwiązania pomocnicze w celu zeskanowania obiektu ze wszystkich stron bez jego obracania (np. zestaw specjalnie ustawionych luster) [16].



Rys. 7. Zasada działania skanera bazującego na świetle strukturalnym

Fig. 7. The principle of operation of the structured light sensor

W metodach opartych na sekwencji wzorców obiekt przesuwający się między kolejnymi naświetleniami powoduje zakłamanie wyników. Dlatego do analizy rzeczywistych, dynamicznie zmiennych scen stosuje się jedynie wzorce pojedyncze. Dzięki temu każda klatka obrazu zawiera informacje o całym modelu. Można tu wyróżnić metody oparte na wzorcach kodowanych geometrycznie i kolorowo. W pierwszej z metod stosowane są jednobarwne wzorce geometryczne zakodowane tak, aby poszczególne jego bloki były unikalne w pewnym otoczeniu. W przypadku drugiej metody stosowane są np. różnokolorowe pasy lub szachownice, a z układu kolorów rekonstruowana jest powierzchnia obiektów (rys. 9). Rozwiązania te charakteryzują się dużą szybkością działania, od kilkunastu do ponad stu klatek przetwarzanych w ciągu sekundy [2].

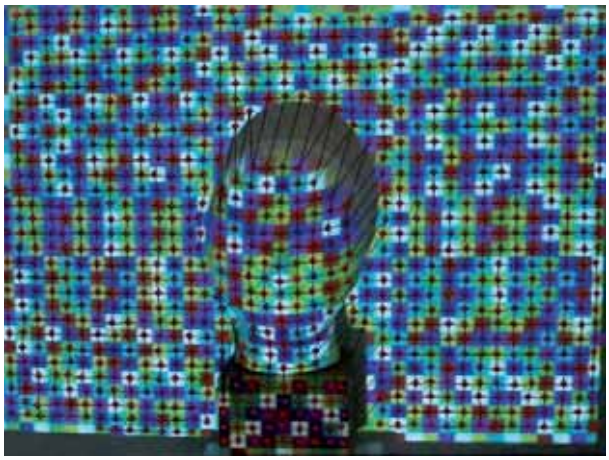
Rys. 8. Przykład wykorzystania światła strukturalnego do modelowania obiektów; od góry – układ pomiarowy; na dole dwa wybrane etapy oświetlania prążkami Graya i wynikowy model otrzymany po oświetleniu 40-toma wzorcami [web.media.mit.edu/~dlanman]

Fig. 8. Example of usage of structured light for object modeling. The first picture presents the measurement system setup, next two present exemplary phases of gray-encoded pattern projection. The final object model is generated based on 40 consecutive pattern illuminations [web.media.mit.edu/~dlanman]

W zastosowaniach robotycznych, gdy maszyny mają działać wśród ludzi, wzorce kodowane kolorami są niewygodne – projektor musi działać w paśmie światła widzialnego, co może przeszkadzać przebywającym w pobliżu ludziom. W takim przypadku zdecydowanie lepiej sprawdzają się wzorce geometryczne, których rzutniki mogą działać w podczerwieni, w sposób niewidoczny i nieprzeszkadzający użytkownikom. W taki sposób działa np. czujnik Kinect firmy Microsoft.

Największą niedogodnością związaną z wykorzystaniem obrazowania opartego na rzutowaniu wzorców jest konieczność dokładnego wykrycia tego wzoru. Dlatego też najczęściej stosowane jest na niewielkie odległości, przy skanowaniu pojedynczych obiektów. Przy stosowaniu na większe odległości konieczne jest stosowanie projekcji w paśmie podczerwonym (aby wykluczyć zakłócenia od

źródeł światła obecnych normalnie w scenie, jak lampy) lub wykorzystanie projektorów o bardzo dużej mocy. Żaden z wariantów nie sprawdza się jednak na otwartej przestrzeni, gdzie światło słoneczne zakłóca działanie praktycznie wszystkich sensorów tego typu.



Rys. 9. Jeden ze sposobów kolorowego kodowania wzorców [2]
Fig. 9. An example of color-encoded structured light pattern [2]

W przypadku stosowania układu złożonego z pojedynczej kamery i projektora występuje silne zjawisko cienia, związane z fizycznym przesunięciem względem siebie obu urządzeń. Można je zminimalizować stosując układ złożony z co najmniej dwóch kamer, umieszczonych po przeciwnych stronach projektora. Problem wyrównania map głębi i obrazu kolorowego nie występuje przy stosowaniu tej samej kamery do akwizycji koloru i głębi, a więc wzorców rzutowanych w paśmie widzialnym (rys. 9), gdzie wyrównanie jest idealne (jak w stereowizji). W przypadku, kiedy do akwizycji obrazu kolorowego stosowana jest inna kamera niż do akwizycji wzorca rzutowanego w podczerwieni, konieczne jest dopasowanie obu obrazów, co wprowadza pewne błędy oraz zajmuje dodatkowy czas.

4.3. Rozwiązania programowe i sprzętowe

W przypadku stosowania algorytmów zaimplementowanych programowo uruchamianych na komputerze sterującym można wymienić praktycznie te same wady, jak przy stereowizji. Największą z nich jest obciążenie systemu, gdyż algorytm jest dość skomplikowany. Rozwiązania sprzętowe rozwiązują problem szybkości działania i obciążenia komputera sterującego, jednak ich cena przez bardzo długi czas była wysoka (rzędu setek do tysięcy USD). Pod koniec 2010 r. pojawił sie na rynku czujnik Kinect, realizujący sprzętowo analizę deformacji wzorca.

Wcześniej w robotyce mobilnej stosowane były głównie rozwiązania bazujące na obrazowaniu na podstawie projekcji pojedynczych linii, głównie ze względu na prostotę koniecznych obliczeń i szybkość działania całego systemu [3, 28].

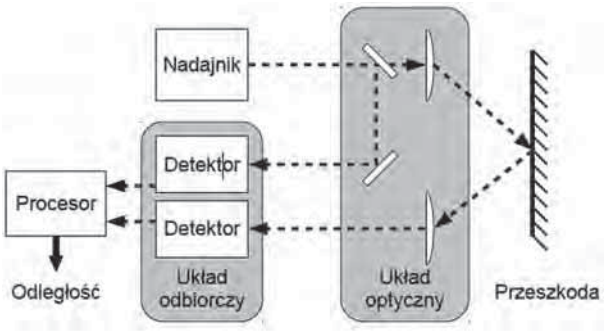
5. Pomiar czasu przelotu sygnału

Ostatnią z trzech głównych metod pomiaru odległości jest pomiar czasu przelotu sygnału ToF (ang. *Time Of Flight*), np. wiązki światła lub ultradźwięków. Istnieje cały szereg różnych czujników działających zgodnie z tą ideą, od najprostszych, jednopunktowych, takich jak np. sonary ultradźwiękowe, dalmierze laserowe, przez planarne czujniki laserowe aż po dwuwymiarowe matryce kamer ToF.

Czujniki opisywanego typu, oprócz informacji o odległości, często zwracają też informację o intensywności odebranej wiązki, która może zostać wykorzystana do stworzenia obrazu I-D, czyli obrazu monochromatycznego z idealnie wyrównaną mapą głębi. Należy jednak pamiętać, że intensywność ta nie musi wcale być dokładnie odpowiednikiem jasności obiektu – zależy ona od współczynnika pochłaniania fal danej długości przez obiekt, która np. w przypadku sonarów ultradźwiękowych jest całkowicie różna od percepcji ludzkiego oka.

5.1. Zasada działania

W przypadku, kiedy mierzona fala ma stosunkowo niską prędkość propagacji w danym ośrodku (np. dźwięk w powietrzu) lub odległości są stosunkowo duże, możliwy jest bezpośredni pomiar czasu potrzebnego na pokonanie przez sygnał drogi od generatora do obiektu i z powrotem do sensora (rys. 10). Wartość ta jest wprost proporcjonalna do odległości od obiektu.



Rys. 10. Schemat ideowy działania kamer ToF mierzących bezpośrednio czas przelotu wiązki światła
Fig. 10. The principle of operation of ToF cameras (direct measurement of the time of flight of a ray)

Czujniki ToF wykorzystujące światło (najczęściej w bliskiej podczerwieni) mogą działać na różne sposoby. Z powodu bardzo dużej szybkości nośnika, dla uzyskania wyników o dobrej rozdzielczości konieczne byłoby zastosowanie bardzo dokładnych układów pomiarowych o dokładności poniżej 1 ns. W takim przypadku światło rzutowane na scenę jest modulowane sinusoidalnie z pewną częstotliwością. Odbiornik przez pewien okres rejestruje w każdym z punktów jasność i na tej podstawie określa fazę odebranego sygnału. Na podstawie porów-

nania jej z fazą sygnału emitowanego można wyznaczyć bezpośrednio odległość, w jakiej znajduje się przedmiot, który to światło odbił.

Wynik będzie prawdziwy tylko w sytuacji, gdy odbita fala jest przesunięta o mniej niż pół okresu (ze względu na okresowość fali nośnej większe przesunięcia są źle wyznaczane). Wynika z tego maksymalny zakres działania czujników, który dla 20 MHz (popularna częstotliwość stosowana we współczesnych czujnikach ToF) daje zakres pomiarowy 7,5 m. Czytelnikom zainteresowanym dokładnym poznaniem techniki działania kamer ToF polecamy pozycję [17].

5.2. Wykorzystanie

W celu uzyskania pełnej mapy głębi przy wykorzystaniu prostszych, jednowymiarowych lub planarnych czujników, konieczne jest ich zamontowanie na ruchomych głowicach i sekwencyjne skanowanie otoczenia. Rozdzielczość tak tworzonych map głębi ograniczona jest dokładnością zastosowanych serwonapędów głowicy (chodzi głównie o możliwą do uzyskania kątowa rozdzielczość ustawianej pozycji). Jeśli jako czujnik zastosuje się skaner laserowy, to jego rozdzielczość kątowa i zakres pomiaru ograniczają efektywną rozdzielczość pomiaru w jednym z kierunków – poziomym lub pionowym (w zależności od sposobu mocowania i osi obrotu). Rozdzielczość zależy też wprost od czasu wymaganego do realizacji pojedynczego, pełnego skanu. Jeśli sceny są statyczne, to skanować można z mniejszą rozdzielczością kątowa, jeśli natomiast wymagany jest szybki skan całego otoczenia, rozdzielczość ta musi być zmniejszona.

Rys. 11. Ruchomy skaner laserowy zamontowany na głowicy uchylnej w szyi robota PR2

Fig. 11. A tilting laser scanner mounted on the PR2 robot

Dokładność uzyskiwanej głębi zależy praktycznie jedynie od zastosowanego czujnika – najsłabsze wyniki uzyskuje się przy stosowaniu czujników ultradźwiękowych lub czujników podczerwieni, które są jednymi z najtańszych możliwych do zastosowania. Na drugim biegunie są skanery laserowe (np. Sick LMS), te jednak są drogie i ciężkie, co wymusza stosowanie droższych i dokładniejszych serwonapędów. Przykładowa integracja skanera laserowego oraz głowicy uchylnej z szyją robota przedstawiona została na rys. 11.

Najprostszym sposobem wykorzystania uzyskiwanej informacji jest zapisanie otrzymanych wyników wprost w mapie głębi. Można jednak wykorzystać fakt, że skany wykonywane są w pewnych określonych sekwencjach, głównie liniami. Dzięki temu w trakcie zbierania kolejnych pomiarów można od razu interpretować i przetwarzać je w celu stworzenia innej reprezentacji otoczenia. W każdej kolejnej zeskanowanej linii można wyszukiwać za pomocą odpowiednich algorytmów odcinki proste, a te zebrane w kolejnych skanach mogą być składane w większe płaszczyzny. Obliczenia te wykonywane są w tym samym czasie co zbieranie pomiarów, dzięki czemu praktycznie od razu po wykonaniu ostatniego skanu dostaje się oprócz samej mapy głębi, dodatkowe informacje o geometrii sceny. Kolejne etapy opisanej agregacji danych pokazano na rys. 12.

Rys. 12. Kolejne poziomy agregacji danych zebranych przy pomocy skanera laserowego; od lewej – chmura punktów, odcinki utworzone z kolejnych odczytów, płaszczyzny określające przeszkody [25]

Fig. 12. Consecutive steps of planar laser scaner measurements aggregation. From left: initial point cloud, aggregated line segments, planar surfaces describing obstacles [25]

Czujniki ToF są szeroko stosowane w robotyce, zarówno w robotach operujących w pomieszczeniach [18], jak i działających w środowisku zewnętrznym [27]. Dokładność otrzymywanych pomiarów, szybkość działania i niskie obciążenie komputera sterującego umożliwiają wykorzystanie otrzymywanego obrazu nie tylko do wykrywania i omijania przeszkód, ale także do budowy mapy otoczenia i samolokalizacji robota (na podstawie obserwacji punktów charakterystycznych w mapie głębi).

5.3. Właściwości metody

W przypadku stosowania czujników ToF zamontowanych na efektorze (np. robocie mobilnym a nawet głowicy) problematyczne może być wyrównanie uzyskanych map głębi z obrazem kolorowym, pozyskiwanym przy użyciu oddzielnego urządzenia. Zazwyczaj występują dość duże niedokładności przy łączeniu obrazu RGB z mapą D, a więc ostateczna jakość obrazu RGB-D jest znacząco niższa niż obrazu I-D uzyskiwanego z samego sensora

ToF. Przy stosowaniu czujników dwuwymiarowych możliwe jest kilka metod rozwiązania problemu wyrównania obrazu kolorowego. Pierwsza z nich to całkowicie niezależna kamera kolorowa umieszczona możliwie blisko kamery ToF. Rozwiązanie to jest proste w realizacji, uzyskiwana mapa RGB-D może być jednak niedokładna. Inną metodą jest umieszczenie obu kamer w jednym urządzeniu i zastosowanie układu optycznego rozdzielającego światło w paśmie widzialnym do kamery RGB, a światło podczerwone kierując do sensora ToF. W tym wypadku obie kamery korzystają fizycznie z tego samego obiektywu i widzą dokładnie to samo, a więc wyrównanie jest niemalże idealne (z dokładnością do jakości montażu układu optycznego i charakterystyki optyki związanej z różnym zniekształceniem fali o innych długościach). Można też zastosować mieszaną matrycę CMOS, zawierającą naprzemiennie rzędy pikseli czułych na kolor (jak w klasycznej kamerze) oraz układów pomiarowych ToF. Takie rozwiązanie pozwala na najlepsze wyrównanie uzyskanych map, jest jednak dość skomplikowane technologicznie.

W każdym wypadku obciążenie systemu wnoszone przez akwizycję danych z sensorów ToF jest minimalne – urządzenia zwracają bezpośrednio odległość do obiektów, w przypadku kamer ToF z częstotliwością nierzadko powyżej 100 FPS. Wadą jest dość niska rozdzielczość matryc sensorów odległości – obecnie produkowane sensory mają rozdzielczość maksymalną rzędu 320 px × 240 px, a tańsze modele często mają rozdzielczość poniżej 100 px × 100 px. Sama technika wykonywania pomiaru ma kilka cech, które mogą utrudniać jej wykorzystanie, np. występowanie wielokrotnych odbić, przez co do sensora może dotrzeć i zostać zarejestrowane światło odbite i załamane od obiektów znajdujących się bliżej niż faktyczny punkt widziany w danym miejscu matrycy. Podobnie zafałszowane wyniki pomiarów mogą wystąpić w obecności silnych źródeł światła w obserwowanej scenie. Z kolei przy jednoczesnym stosowaniu wielu kamer ToF należy uwzględnić problem interferencji emitowanego światła – najczęściej rozwiązywany przez sekwencyjne odczyty z kolejnych kamer (co z kolei zmniejsza faktyczną szybkość akwizycji danych).

6. Podsumowanie

W artykule skupiono uwagę na czujnikach zwracających obrazy RGB-D. Wyjaśniono podstawowe terminy oraz rodzaje reprezentacji głębi. Omówiono trzy główne paradygmaty działania czujników, tj. stereowizję – metodę pasywną opartą na jednoczesnym wykorzystaniu pary kamer, oraz dwie metody aktywne – światło strukturalne oraz pomiar czasu lotu wiązki. W drugiej części artykułu zaprezentowane zostaną obecne na rynku czujniki zwracające obrazy RGB-D wykorzystujące omówione techniki.

Podziękowania

Praca finansowana ze środków Narodowego Centrum Nauki, grant 2012/05/D/ST6/03097.

Zaproszenie

Osoby zainteresowane tematyką poruszaną w artykule zapraszamy do udziału w sesji specjalnej Percepcja robotów z wykorzystaniem obrazów RGB-D, organizowanej w ramach 13. Krajowej Konferencji Robotyki – Kudowa Zdrój, 2-6 lipca 2014 r. www.kkr13.pwr.wroc.pl.

Bibliografia

1. Beetz M., Burgard W., Cremers D., Pangercic D., Sturm J., RGB-D Workshop on 3D Perception in Robotics. *Part of the European Robotics Forum*, 2011.
2. Chen S., Li Y., Zhang J., *Vision processing for realtime 3-d data acquisition based on coded structured light*, *Image Processing, IEEE Transactions on*, 17(2), 2008, 167–176.
3. Evans J., Krishnamurthy B., Barrows B., Skewis T., Lumelsky V., Handling real-world motion planning: a hospital transport robot. *Control Systems, IEEE*, 12(1):15–19, feb 1992. 4. A. Fossati, J. Gall, H. Grabner, M. Hansard. 3rd IEEE Workshop on Consumer Depth Cameras for Computer Vision. *Workshop in conjunction with International Conference on Computer Vision* (ICCV), 2013.
4. Fossati A., Gall J., Grabner H., Ren X., Konolige K., 1st Workshop on Consumer Depth Cameras for Computer Vision. *Workshop in conjunction with 13th International Conference on Computer Vision* (ICCV), 2011.
5. Fossati A., Gall J., Grabner H., Ren X., Konolige K., Lee S., Hansard M., 2nd Workshop on Consumer Depth Cameras for Computer Vision. *Workshop in conjunction with 12th European Conference on Computer Vision* (ECCV), 2012.
6. Fox D., Konolige K., Kosecka J., Ren X., RGB-D: Advanced Reasoning with Depth Cameras. *Workshop in conjunction with Robotics: Science and Systems* (RSS), 2010.
7. Fox D., Konolige K., Kosecka J., Ren X.. RGB-D: Advanced Reasoning with Depth Cameras. *Workshop in conjunction with Robotics: Science and Systems* (RSS), 2011.
8. Fox D., Konolige K., Kosecka J., Ren X.. RGB-D: Advanced Reasoning with Depth Cameras. *Workshop in conjunction with Robotics: Science and Systems* (RSS), 2012.
9. Godin G., Goesele M., Matsushita Y., Sagawa R., Yang R. (eds) *Special Issue on 3D Imaging, Processing and Modeling Techniques*, wol. 102, *International Journal of Computer Vision*. IEEE, Mar. 2013.
10. Han C., Jiang Z., Indexing coded stripe patterns based on de bruijn in color structured light system. *National Conference on Information Technology and Computer Science* (CITCS), 621–624, 2012.
11. Hirschmuller H., Stereo processing by semiglobal matching and mutual information, *IEEE Trans. Pattern Anal. Mach. Intell.*, II 2008, 328–341.
12. Konolige K., Projected texture stereo. *International Conference on Robotics and Automation* (ICRA), IEEE, 2010, 148–155.

13. Konolige K., Agrawal M., Bolles R.C., Cowan C., Fischler M.A., Gerkey B.P., Outdoor mapping and navigation using stereo vision, *International Symposium on Experimental Robotics*, 2006, 179–190.
14. Kornuta T., Percepcja robotów z wykorzystaniem obrazów RGB-D. *Sesja specjalna 13. Krajowej Konferencji Robotyki* (KKR), 2014.
15. Lanman D., Crispell D., Taubin G.. Surround structured lighting for full object scanning. *3-D Digital Imaging and Modeling, 3DIM'07. 6th International Conference on*, IEEE, 2007, 107–116.
16. Lee S., Choi O., Horaud R., *Time-of-flight cameras: principles, methods and applications*. Springer, 2013.
17. Prusak A., Melnychuk O., Roth H., Schiller I., Koch R., Pose estimation and map building with a Time-Of-Flight-camera for robot navigation. *Int. J. Intell. Syst. Technol. Appl.*, 5:355–364, November 2008.
18. Ribo M., Brandner M., State of the art on vision-based structured light systems for 3D measurements. *International Workshop on Robotic Sensors: Robotic and Sensor Environments,* 2–6, 2005.
19. Rusu R.B., Aldoma A., Gedikli S., Dixon M., 3D Point Cloud Processing: PCL. *Tutorial at IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS), 2011.
20. Salvi J., Pages J., Batlle J., Pattern codification strategies in structured light systems, *Pattern Recognition*, 37(4):827–849, 2004.
21. Saxena A., Koppula H., Newcombe R., Ren X.. RGB-D: Advanced Reasoning with Depth Cameras. *Workshop in conjunction with Robotics: Science and Systems* (RSS), 2013.
22. Shao L., Han J., Xu D., Shotton J. (eds.), *Special issue on Computer Vision for RGB-D Sensors: Kinect and Its Applications*, vol. 43, *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernectics*, 2013.
23. Stefańczyk M., Kasprzak W., Multimodal segmentation of dense depth maps and associated color information. *Proceedings of the International Conference on Computer Vision and Graphics*, vol. 7594, *Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, 2012, 626–632.
24. Surmann H., Lingemann K., Nüchter A., Hertzberg J., A 3d laser range finder for autonomous mobile robots. *32nd International Symposium on Robotics* (ISR), 2001, 153–158.
25. Tao T., Koo J.C., Choi H. R., A fast block matching algorthim for stereo correspondence, *IEEE Conference on Cybernetics and Intelligent Systems*, 38–41, 2008.
26. Thrun S., Montemerlo M., Dahlkamp H., Stavens D., Aron A., Diebel J., Fong P., Gale J., Halpenny M., Hoffmann G., Lau K., Oakley C., Palatucci M., Pratt V., Stang P., Stanley: The robot that won the DARPA Grand Challenge. *Journal of Field Robotics*, 23(9), 2006, 661–692.
27. Wei B., Fan Y., Gao B.. Mobile robot vision system based on linear structured light and DSP. *International Conference on Mechatronics and Automation,* ICMA 2009, 1285–1290.

Acquisition of RGB-D images: methods

Abstract: The two-part article is devoted to sensors enabling the acquisition of depth information from the environment. The following, first part concentrates on three main methods of depth measurement: stereovision, structured light and time of flight (ToF). Along with the principle of operation of each of the method we also deliberate on their properties, analyse the complexity of required computations and present potential applications.

Keywords: depth map, point cloud, RGB-D image, RGB-D sensor, time-of-flight, structured light, stereovision

Artykuł recenzowany, nadesłany 02.11.2013 r., przyjęty do druku 20.12.2013 r.

mgr inż. Maciej Stefańczyk

Absolwent Wydziału Elektroniki i Technik Informacyjnych Politechniki Warszawskiej. W 2010 r. uzyskał tytuł inżyniera, w 2011 r. tytuł magistra inżyniera, obydwa z wyróżnieniem. W 2011 r. rozpoczął prace nad doktoratem dotyczącym zastosowania aktywnej wizji wraz z systemami opartymi na bazie wiedzy w sterowaniu robotów. Główne zainteresowania naukowe obejmują zastosowanie informacji wizyjnej, zarówno w robotyce, jak i w systemach rozrywki komputerowej.

e-mail: stefanczyk.maciek@gmail.com

dr inż. Tomasz Kornuta

Absolwent Wydziału Elektroniki i Technik Informacyjnych Politechniki Warszawskiej. W 2003 r. uzyskał tytuł inżyniera, w 2005 r. tytuł magistra inżyniera, a w 2013 r. stopień doktora nauk technicznych. Od 2008 r. pracuje w Instytucie Automatyki i Informatyki Stosowanej, a od 2009 r. pełni funkcję kierownika Laboratorium Podstaw Robotyki. Jego zainteresowania naukowe obejmują metody programowania robotów oraz wykorzystanie informacji wizyjnej w robotyce, a w szczególności aktywną wizję oraz rozpoznawanie obrazów RGB-D. Autor/współautor ponad trzydziestu publikacji dotyczących ww. tematów. Recenzent krajowych oraz międzynarodowych konferencji robotycznych (KKR, IEEE MMAR, IEEE ICAR, IFAC SYROCO) oraz czasopism (Sensor Review, International Journal of Advanced Robotics). Członek IEEE RAS.

e-mail: tkornuta@ia.pw.edu.pl