

Wykrywanie zasięgu

- * Opisz problemy z odbiciami zwierciadlanymi, przesłuchami i skrótami perspektywicznymi za pomocą przetwornika ultradźwiękowego, a jeśli otrzymamy rysunek liniowy powierzchni 2D, zilustruj, gdzie każdy z tych problemów może wystąpić.
- * Opisz reprezentację wiedzy o chmurze punktów i rodzaje algorytmów z nią związanych.
- * Opisz dwa popularne typy czujników aktywnego zasięgu do wytwarzania chmur punktów: kamery lidarowe i RGB-D.
- * Zapoznaj się z sekwencją Sense-Register-Reconstruct do przetwarzania chmur punktów.

Przegląd

Aby inteligentnie funkcjonować na świecie, roboty muszą być w stanie mierzyć lub wyodrębnić głębokość lub zasięg. Prowadzi to do pytania: Jaka jest różnica między głębokością a zasięgiem? Terminy „głębokość” i „zakres” są używane w robotyce jako synonimy w odniesieniu do odległości między robotem a powierzchnią na świecie. Jednak społeczność wizji komputerowych często używa „głębokości” w odniesieniu do niewielkich zmian odległości, które wskazują boki obiektów w celu rekonstrukcji powierzchni i szacowania wolumetrycznego obiektu. Czujniki o wysokiej rozdzielczości, które można wykorzystać do określenia trójwymiarowego kształtu obiektu, są ogólnie nazywane czujnikami głębokości, aby odróżnić te czujniki od czujników zasięgu. Społeczność robotyki zazwyczaj używa zasięgu jako miary odległości do przeszkody lub wzniesienia terenu; pomiary z czujników odległości mogą mieć znacznie niższą rozdzielczość do nawigacji niż do manipulacji obiektami. W tym rozdziale znajduje się również odpowiedź: Dlaczego robotom tak trudno jest wyczuć głębokość i zasięg? Zasięg można wykrywać za pomocą czujników pasywnych lub aktywnych. Czujnik pasywnego zasięgu to zazwyczaj pojedyncza kamera lub para kamer stereo. Jednym z podejść jest wykorzystanie przepływu optycznego, jak opisano wcześniej. Innym podejściem do obliczania zasięgu jest wyodrębnienie głębi z obrazu na podstawie triangulacji cech widocznych na dwóch obrazach lub ze wskazówek w magu. Przy takim podejściu wyzwaniem jest znalezienie na obrazach odpowiadających sobie cech, zwykle za pomocą algorytmów zwanych operatorami zainteresowań. Najpopularniejszym algorytmem wykrywania odpowiednich cech jest transformacja cech niezmienna skali (SIFT). Na szczęście istnieje wiele wskazówek dotyczących odległości i głębokości, na przykład względny rozmiar obiektu, które tworzą pole widzenia komputera znane jako głębia od X , gdzie X jest używanym wskaźnikiem. Aktywne czujniki generalnie „malują” obszar przestrzeni jakimś rodzajem emisji. Aktywny czujnik może albo bezpośrednio zmierzyć zasięg, albo wydobyć głębokość. Aktywne czujniki, takie jak sonary i lidar, wysyłają odpowiednio impuls dźwiękowy lub świetlny do otoczenia i mierzą czas lotu, aby sygnał powrócił. Ponieważ prędkość emisji energii jest znana, zasięg do obiektu wynosi $\frac{1}{2} \text{time}_{\text{flight}} \times \text{speed}$. Aktywny czujnik może również pomalować obszar światłem i wykorzystać wzór światła jako cechy, które można zaobserwować na pojedynczym obrazie. Eliminuje to potrzebę posiadania wielu obrazów, stosowania operatorów zainteresowań w celu dopasowania funkcji, a następnie wyodrębniania z nich funkcji głębi wykorzystujące stereo lub inną głębię z algorytmów X . Historycznie rzecz biorąc, roboty po raz pierwszy badali kamery i pary kamer RGB, począwszy od późnych lat 60. XX wieku, naśladując widzenie biologiczne, ale koszty obliczeniowe algorytmów przekraczały możliwości komputerów w tamtych czasach i było niewiele przekonujących sukcesów. Czujniki zasięgu laserowego były opracowywane dla Departamentu Obrony, ale rozmiar i koszt (ponad 100 000 USD) były zaporowe. Na szczęście dokoptowano niedrogi czujniki ultradźwiękowe (30 USD) używane do autofokusa aparatu Polaroid. Czujniki miały bardzo niską rozdzielczość przestrzenną i były hałaśliwe, ale były tanie i na tyle małe, że robot mógł unieść ich szereg lub pierścień. Zbadano szereg algorytmów probabilistycznych i reprezentacji wiedzy, aby spróbować

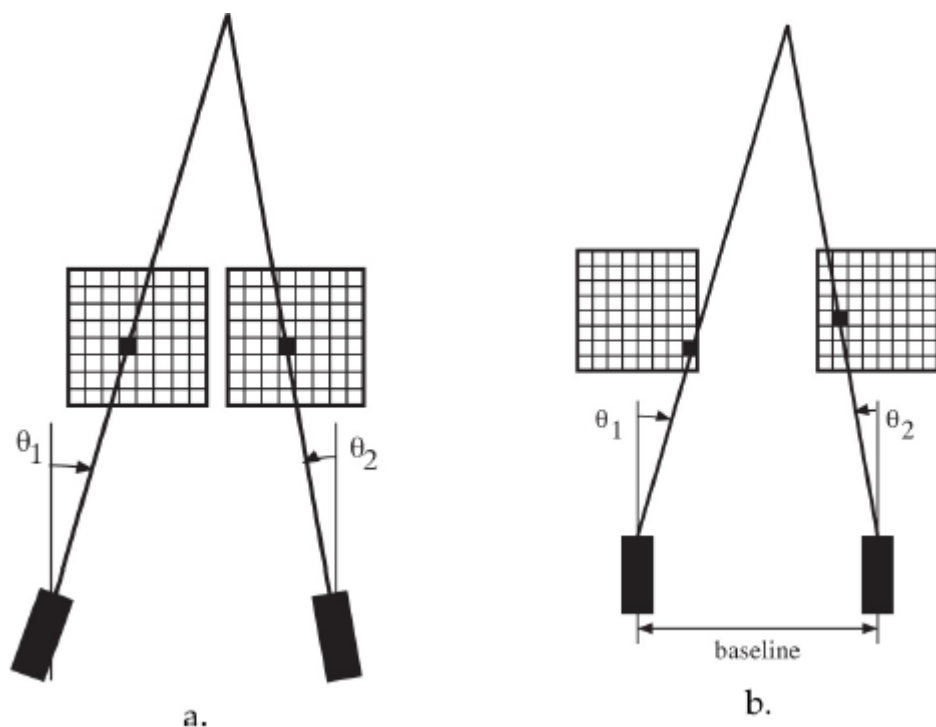
wykonać wiele odczytów o niskiej rozdzielczości i tworzyć mapy zasięgu w wysokiej rozdzielczości. Metody te przyczyniły się do opracowania przestrzennej reprezentacji siatki zajętości i metod wnioskowania dowodowego, które później stały się podstawą algorytmów jednoczesnej lokalizacji i mapowania (SLAM) omówionych w rozdziale 15. Pole to przeniosło się na czujniki zasięgu laserowego, gdy SICK, niemiecka firma specjalizująca się w czujnikach, wprowadziła bardzo dokładny dalmierz laserowy o wysokiej rozdzielczości do pomiarów geodezyjnych, który generował mapę zasięgu w płaszczyźnie poziomej we współrzędnych cylindrycznych. W końcu spadły koszty i rozmiary laserowych tropicieli z trójwymiarowymi mapami zasięgu, a tropiciele, zwłaszcza czujnik Velodyne, stały się powszechne w robotach zewnętrznych. Wprowadzenie egocentrycznych informacji o zasięgu o wysokiej rozdzielczości doprowadziło do powstania nowej struktury wiedzy, chmury punktów. Chmura punktów może być odwzorowana na bezwzględny układ współrzędnych, w którym każdy region przestrzeni 3D jest oznaczony jako zajęty lub pusty. Mapa jest zasadniczo siatką zajętości, ale teraz dominuje termin chmura punktów. Być może największą zmianą w zakresie wykrywania zasięgu na przestrzeni lat był rozwój systemów RGB-D, w których D oznacza głębię, takich jak Kinect firmy Microsoft, który wykorzystywał niedrogie czujniki aktywne i pasywne z wyrafinowanymi algorytmami stereo, które można teraz przetwarzać obliczeniowo na nowoczesnych procesorach. Bez względu na to, jak wielkim przełomem był Kinect, odpowiedź na pytanie, czy Microsoft Kinect nie rozwiązuje wszystkiego? jest nie, ponieważ system jest wrażliwy na warunki środowiskowe i działa tylko na krótkich dystansach. Projekt Google Tango® to kolejny przykład czujników RGB-D. Chociaż nie pojawiły się konsumenckie zastosowania takich czujników, czujniki RGB-D wywarły duży wpływ na robotykę. W tym rozdziale omówiono główne klasy algorytmów i reprezentacji wiedzy do określania zasięgu obiektu lub powierzchni. Zaczyna się od pasywnych par stereo wizyjnych, po których następuje aktywna ultradźwięki. Chociaż ultradźwięki są rzadko używane ze względu na ich znaczące wady, które zostaną omówione w dalszej części tej części, doprowadziły one do reprezentacji wiedzy o siatce zajętości i wykorzystania rozumowania dowodowego do budowania pewności na podstawie zaszumionych odczytów. Ta spuścizna przetrwała do dziś. Wizja komputerowa poszukiwała mniej skomplikowanych obliczeniowo alternatyw dla par stereo, w szczególności głębi z algorytmów X. Inną próbą niedrogiego aktywnego wykrywania były lekkie paski, które rzucają światło na scenę i wywnioskują głębię na podstawie zniekształcenia w linii prostej. Lidar nas najbardziej skuteczny i dokładny czujnik do bezpośredniego pomiaru zakresu, ale pozostaje najdroższy. Kamery RGB-D to przeformułowanie jasnych pasków. Zarówno kamery lidarowe, jak i RGB-D generują niezwykle gęste zestawy danych, które stanowią wyzwanie dla złożoności obliczeniowej; te zestawy danych są reprezentowane jako chmury punktów, które wymagają dodatkowego przetwarzania w celu zarejestrowania z innymi czujnikami i wyodrębnienia powierzchni.

Stereo

STEREOPSISIS

PARA STEREO

Ludzie wydobywają głębię ze wzroku. W większości przypadków, choć nie zawsze, percepcja głębi wynika z posiadania dwojga oczu i zdolności do triangulacji, jak pokazano na rycinie



, znanej również jako stereopsja. Korzystanie z dwóch kamer w celu wyodrębnienia danych o zasięgu jest często określane jako zakres od stereo, rozbieżności stereo, widzenia obuocznego lub po prostu "stereofoniczny." Jednym ze sposobów na wydobycie głębi jest próba nałożenia kamery na każde oko, jak na rysunku a. Każda kamera znajduje ten sam punkt na każdym obrazie, obraca się do środka tego punktu na obrazie, a następnie mierzony jest kąt względny. Kamery są znane jako para stereo.

KORESPONDENCJA

ZAINTERESOWANY OPERATOR

Ta metoda ma dwa wyzwania. Po pierwsze, trudno jest zaprojektować i zbudować mechanizm, który może precyzyjnie poruszać się, aby zbiegać się w punktach. (Jeszcze trudniej jest zaprojektować i zbudować niedrogi mechanizm wergencji). Drugie wyzwanie jest jeszcze bardziej fundamentalne i rodzi pytanie: Skąd robot wie, że na obu obrazach patrzy w ten sam punkt? Wyzwanie to jest określane jako problem korespondencji, ponieważ zadaniem jest znalezienie punktu na jednym obrazie, który odpowiada punktowi na drugim obrazie. Typowym podejściem jest identyfikacja „interesujących” lub potencjalnie wyjątkowo cennych pikseli na obrazie, takich jak bardzo jasne lub ciemne punkty lub krawędzie. Algorytm, który wybiera interesujące piksele, nazywa się operatorem zainteresowań. Ponieważ nawet najmniejsze zmiany w oświetleniu mają wpływ na obraz, nie ma gwarancji, że oba obrazy, nawet uzyskane dokładnie w tym samym czasie z dwóch kamer, „widzą” te same wartości dla odpowiednich pikseli. Dlatego algorytmy operatora zainteresowania zwykle zwracają listę interesujących pikseli, a nie tylko jeden, a algorytm dopasowujący próbuje znaleźć najlepszą zgodność między nimi wszystkimi. Po ustaleniu punktów zainteresowania, pozostałe piksele muszą być oznaczone głębokością w stosunku do tych punktów.

OBRAZY REKTYFIKOWANE

RÓŻNICA

Na szczęście nie jest konieczne posiadanie mechanicznego systemu wergencyjnego. Zamiast tego, kamery mogą być montowane w miejscu z osiami optycznymi równoległymi do siebie i prostopadłymi

do mocowania, tworząc obrazy rektyfikowane.⁵⁵ Ten typ tradycyjnej „głowy” stereo pokazano na rysunku .



Przestrzeń między osiami dwóch kamer nazywana jest linią bazową. Odległość w miejscu interesującego punktu między obrazami nazywana jest rozbieżnością; odległość punktu od kamer jest odwrotnie proporcjonalna do dysproporcji.

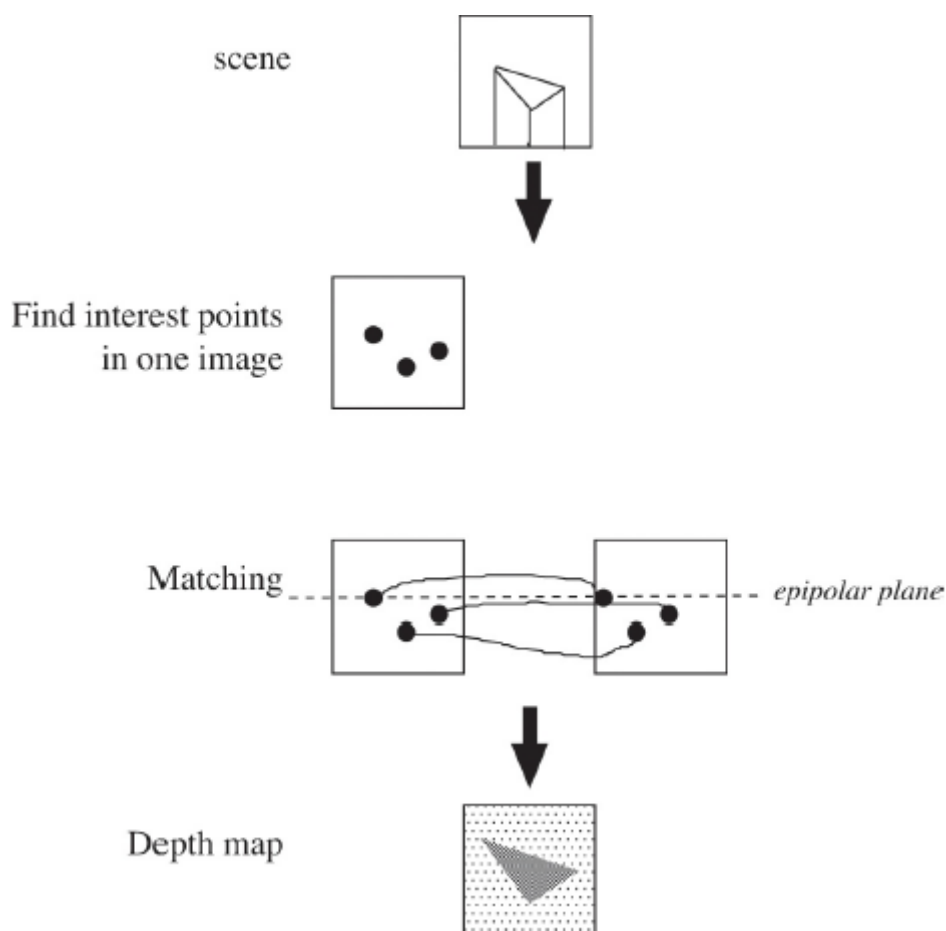
LINIE EIPOLARNE

Podczas gdy obrazy rektyfikowane eliminują potrzebę stosowania drogich mechanicznych systemów wergencyjnych, nie rozwiązują problemu korespondencji. Jeśli kamery są dokładnie wyrównane, jeden wiersz na lewym obrazie będzie odpowiadał wierszowi na prawym obrazie. Mówi się, że te rzędy są liniami epipolarnymi lub rzutami na płaszczyznę epipolarną. Ilekroć robot znajdzie interesujący punkt na jednym obrazie, musi brać pod uwagę tylko piksele wzdłuż epipolarnej linii na drugim obrazie. To ogromne oszczędności obliczeniowe. Działa to jednak tylko wtedy, gdy kamery są idealnie dopasowane optycznie i pozostają w jednej linii. W praktyce roboty poruszają się, podskakują i cierpią z powodu dryftów wyrównania, a kamery mogą mieć pewne wady optyki. Wyrównanie może być okresowo kompensowane w oprogramowaniu poprzez proces kalibracji kamery, w którym robot jest prezentowany ze standardem, a następnie tworzy tabelę lub funkcję kalibracji. W rezultacie wielu badaczy zwraca się do jednostek, które zawierają parę stereofoniczną w jednym stałym przypadku, w którym wyrównanie nie może zostać zmienione. Pierwszym robotem, który z powodzeniem wykorzystał widzenie stereo, był Wózek Stanforda Hansa Moraveca. Moravec pracował nad wózkiem podczas studiów podyplomowych w Stanford w latach 1973-1980. Dłuższe linie bazowe są zwykle bardziej dokładne, ponieważ niewielki błąd pomiaru ma mniejszy wpływ, ale mniejsze linie bazowe mają mniejszy „obszar” i w efekcie zajmują mniej miejsca. Ten sam punkt na obu obrazach wciąż musi zostać zidentyfikowany.

ZAKRES OBRAZU

MAPA GŁĘBOKOŚCI

Rysunek



przedstawia uproszczony przebieg operacji przy wyodrębnianiu zakresu z pary obrazów. Proces rozpoczyna się od dwóch obrazów, pary lewy-prawy, i prowadzi do trzeciego obrazu zwanego obrazem zasięgu lub mapą głębi. Para lewa-prawa może być w skali szarości lub w kolorze, ale mapa głębi jest mapą w skali szarości, w której intensywność jest proporcjonalna do odległości obiektu od kamer. Główną wadą wydobywania informacji o zasięgu z wizji jest to, że algorytmy są zwykle złożone obliczeniowo. Algorytmy dopasowywania stereo są zwykle wykonywane w kolejności $O(n^2m^2)$. Oznacza to, że obraz o rozmiarze 640 na 480 przyjmuje do przetworzenia rzędu 9×10^{10} instrukcji, podczas gdy liczba instrukcji wymaganych do segmentacji koloru jest rzędu $O(nm)$ lub 3×10^5 . Nawet przy zaawansowaniu mikroprocesorów, obliczenie mapy zakresu stereo może zająć kilka minut.

Głębokość od X

Widzenie komputerowe od dawna wykorzystuje proste wskazówki dotyczące głębokości i orientacji powierzchni. Te wskazówki obejmują lokalizację cieni lub cieni, ogniskową, przy której obiektyw autofokusa zbiega się na obiekcie oraz zmiany tekstury. Wskazówki są często określane jako „głębina z cieniowania”, „głębina z ostrości”, „głębina z tekstury” itd., co prowadzi do przydomka „głębina z X”. Polly jest przykładem robota, który potrafi obliczyć głębokość na podstawie tekstury. Służył jako autonomiczny przewodnik wycieczek w MIT AI Laboratory i Brown University na początku lat 90-tych. W tamtych czasach przetwarzanie obrazu było powolne i kosztowne, co było całkowicie sprzeczne z wysokimi częstotliwościami aktualizacji potrzebnymi do nawigacji przez reaktywny robot mobilny. Postrzeganie zachowania polegającego na unikaniu przeszkód opierało się na sprytniej afordancji: teksturze. Korytarze AI Lab były w całości pokryte tą samą wykładziną. „Kolor” dywanu na zdjęciu miał tendencję do zmiany pod wpływem oświetlenia, ale ogólna tekstura lub „ziarno” nie. W tym przypadku teksturę mierzono jako krawędzie na jednostkę powierzchni.

MAPA GŁĘBOKOŚCI PROMIENIOWEJ

Robot podzielił pole widzenia na kąty lub sektory, tworząc radialną mapę głębi lub odpowiednik wykresu biegunowego. Każdy sektor, który wyprodukował miarę tekstury zbliżoną do tekstury dywanu, był traktowany jako pusty. Gdyby osoba stała na dywanie, ta plama miałaby inną teksturę, a robot oznaczałby obszar jako zajęty. Chociaż ta metodologia miała pewne problemy – na przykład silne cienie na podłodze tworzyły „zajęte” obszary – była szybka i elegancka.

Sonar lub ultradźwięki

Sonar odnosi się do każdego systemu wykorzystującego dźwięk do zakresu pomiarowego. Sonary do różnych zastosowań działają na różnych częstotliwościach; na przykład sonar do pojazdów podwodnych używałby częstotliwości odpowiedniej do poruszania się po wodzie, podczas gdy pojazd naziemny używałby częstotliwości bardziej odpowiedniej do podróżowania w powietrzu. Sonary do pojazdów naziemnych wykorzystują częstotliwość ultradźwiękową, tuż na granicy ludzkiego słuchu. W rezultacie terminy „sonar” i „ultradźwięki” są używane zamiennie przy omawianiu wydobywania zasięgu z energii akustycznej.

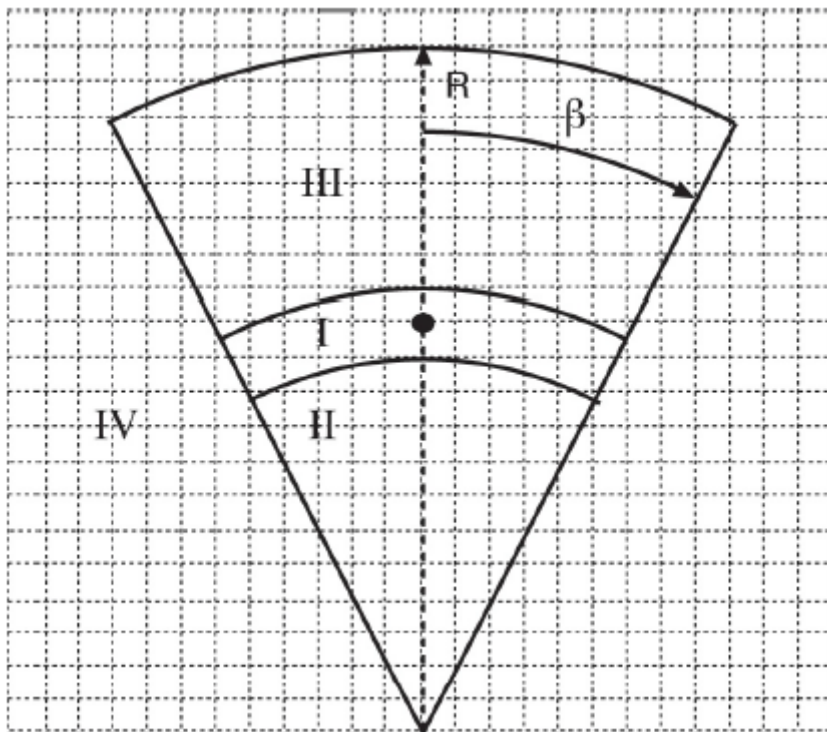
CZAS LOTU

Ultradźwięki to aktywne czujniki, które emitują dźwięk i mierzą czas potrzebny na odbicie dźwięku. Czas lotu (czas od emisji do odbicia) wraz z prędkością dźwięku w tym środowisku (pamiętajmy, że nawet powietrze zmienia gęstość wraz z wysokością) jest wystarczający do obliczenia zasięgu obiektu.

WYKRES BIEGUNOWY

Ultradźwięki były niezwykle powszechne w latach 80. i 90. z kilku powodów. Ich ewolucja zbiegła się z powstaniem paradygmatu reaktywnego. W połowie lat 80. Hans Moravec wykonał imponującą nawigację robotem z pierścieniem sonarów. Konfiguracja pierścienia dawała pokrycie 360° jako wykres biegunowy. Ten pierścień został opracowany przez jednego z pierwszych producentów robotów mobilnych, firmę Denning Robotics. Od tego czasu pierścienie sonaru są często nazywane „pierścieniami Denninga”, niezależnie od producenta. Oprócz zapewniania bezpośrednich pomiarów zasięgu, przetworniki sonaru w pierścieniach były tanie, szybkie i miały wspaniałe zasięgi. Na początku lat 80. firma Polaroid Corporation opracowała małe, niedrogi sonary do użytku jako dalmierze kamer. Większa wersja, przetwornik ultradźwiękowy klasy laboratoryjnej Polaroid, kosztuje rzędu 30 USD i może mierzyć zakresy od 1 do 25 stóp z calową rozdzielczością w polu widzenia 30°. Ponadto czas pomiaru był rzędu sekund w stosunku do godzin w przypadku widzenia komputerowego. Ultradźwięki stały się czujnikiem wybieranym przez roboty oparte na zachowaniu. Przetwornik ma wielkość i grubość monety jednodolarowej i składa się z cienkiej metalowej membrany. Bardzo silny impuls elektryczny przechodzący przez przetwornik generuje kształt fali, powodując, że membrana przetwornika wytwarza dźwięk. Dźwięk jest słyszalny jako cichy trzask, jak krab otwierający i zamykający szczypce. W międzyczasie ustawiony jest zegar i membrana staje się nieruchoma. Odbity dźwięk lub echo wibruje membranę. Wibracje są wzmacniane, a następnie ograniczane przez siłę sygnału zwrotnego; jeśli odbierany jest zbyt mało dźwięku, czujnik przyjmuje, że dźwięk jest szumem i ignoruje go. Jeśli sygnał jest wystarczająco silny, aby osiągnąć próg, zegar zostaje wyzwolony, dając czas lotu. Ustalenie, czy dane sonaru z określonego środowiska będą przydatne, wymaga zrozumienia, w jaki sposób fala dźwiękowa jest generowana przez przetwornik. W rzeczywistości wiązka dźwiękowa wytwarza wiele wtórnych fal dźwiękowych wokół przetwornika, które przed rozproszeniem oddziałują na różne obszary przestrzeni. Wtórne fale dźwiękowe nazywane są płacami bocznymi. Większość systemów robotów zakłada, że dźwięk z głównego lub środkowego płata jest wyłącznie odpowiedzialny za pomiar zasięgu. Szerokość głównego płata jest często modelowana jako 30° szerokości z odległości

około pięciu metrów. Jednak w praktyce roboty reaktywne muszą reagować na przeszkody w zakresie od 0,3 do 3 metrów. W rezultacie wiele algorytmów traktuje płat jako szeroki od 8° do 15° , w zależności od tego, jak wiarygodne są odczyty zakresu w określonym środowisku. W rozdziale 15 bardziej szczegółowo omówimy niezawodność odczytów zakresu. Siła głównego płata w środowisku określa maksymalny zasięg, jaki sonar może wydobyć z niezawodności. W idealnych pomieszczeniach zamkniętych sonar może zwracać zasięg do 25 stóp, podczas gdy na zewnątrz ten sam sonar może dotrzeć do nie więcej niż osiem stóp z dowolną powtarzalnością. Górna granica odczytu zakresu zależy od czujnika i otoczenia; dolna granica nie. Przetworniki ultradźwiękowe mają „czas martwy” bezpośrednio po emisji, podczas gdy vibracje membrany zanikają. Czas zanikania przekłada się na niezdolność do wykrywania obiektów w promieniu 11 cali, ponieważ pomiary wykonane w tym okresie są niewiarygodne, ponieważ membrana mogła nie przestać dzwonić. Niezależnie od maksymalnego dozwolonego powrotu zasięgu (tj. czy program ignoruje odczyty większe niż trzy metry?) i szerokości płata, większość programów komputerowych dzieli obszar objęty sonarem na trzy regiony pokazane na rysunku .



Region I jest powiązany z odczytem zasięgu. Jest to łuk, ponieważ obiekt, który zwrócił dźwięk, może znajdować się w dowolnym miejscu wiązki. Łuk ma szerokość, ponieważ występują pewne błędy rozdzielczości i pomiaru; szerokość Regionu I to tolerancja. Region II to obszar, który jest pusty. Gdyby ten obszar nie był pusty, odczyt zasięgu byłby krótszy. Region III to obszar, który teoretycznie jest objęty wiązką sonaru, ale nie wiadomo, czy jest zajęty, czy pusty, ponieważ znajduje się w cieniu tego, co znajdowało się w regionie I. Region IV znajduje się poza wiązką i nie jest interesujący.

ODBICIE LUSTRZANE

Choć są niedrogie, szybkie i mają duży zasięg działania, ultradźwięki mają wiele mankamentów i ograniczeń, o których konstruktor powinien wiedzieć. Czujniki ultradźwiękowe opierają się na odbiciu, dlatego są podatne na odbicia zwierciadlane. Odbicie zwierciadlane występuje, gdy kształt fali uderza w powierzchnię pod ostrym kątem, a fala odbija się od przetwornika. W idealnym przypadku wszystkie objekty powinny mieć płaską powierzchnię prostopadłą do przetwornika, ale oczywiście rzadko się to

zdarza. Co gorsza, odbity sygnał może odbijać się od drugiego obiektu i tak dalej, aż przez przypadek zwróci on część energii przetwornikowi. W takim przypadku czas lotu nie będzie odpowiadał prawdziwemu względnemu zakresowi.

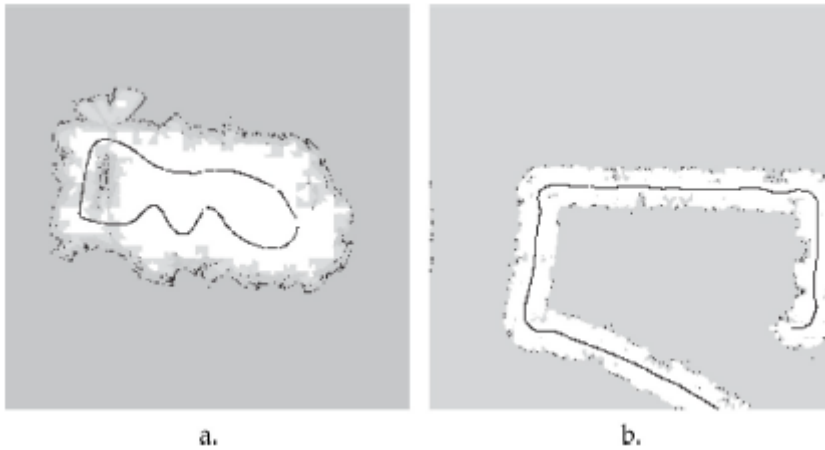
SKRÓT PERSPEKTYWICZNY

Nawet pod bardzo ostrymi kątami powierzchnia jest zwykle wystarczająco szorstka, aby odesłać pewną ilość energii dźwięku. Wyjątkiem jest szkło, które powoduje poważne odbicia zwierciadlane, co jest powszechnym problemem w szpitalach i biurach, w których działają roboty. Na szczęście część energii może nie zostać rozproszona i często jest wystarczająco silna, aby przejść przez progowanie w obwodzie przetwornika. Może jednak pojawić się nowy problem – skrót perspektywiczny. Przypomnij sobie, że sonar ma pole 30° poglądu. Oznacza to, że dźwięk nadawany jest w stożku o szerokości 30° . Jeśli powierzchnia odbijająca nie jest prostopadła do przetwornika, jedna strona stożka dotrze do obiektu pierwsza i najpierw zwróci zasięg. Większość oprogramowania zakłada, że zwrócony odczyt znajduje się wzdłuż osi fali dźwiękowej. Jeśli oprogramowanie korzysta z odczytu (który w rzeczywistości jest odczytem dla 15°), robot zareaguje na błędne dane. Nie ma rozwiązania tego problemu.

PRZESŁUCH

Odbicie zwierciadlane jest nie tylko istotnym źródłem błędnych odczytów, ale może wprowadzić nowy rodzaj błędów w pierścieniach sonarów lub tablicach. Rozważ pierścień z wielu sonarów o zasięgu 360° wokół robota. Załóżmy, że sonary strzelają (wydają dźwięk) mniej więcej w tym samym czasie. Mimo że każdy z nich obejmuje inny obszar wokół robota, niektóre zwierciadlane odbite dźwięki z sonaru mogą zostać odebrane przez zupełnie inny sonar. Echosonda odbiorcza nie jest w stanie odróżnić dźwięku generowanego przez siebie od dźwięku generowanego przez jego rówieśników. Ten błędny odczyt nazywa się przesłuchem, ponieważ fale dźwiękowe się krzyżują. Większość zrobotyzowanych systemów uruchamia odpalenie sonarów naprzemiennie, na przykład w pierścieniu Denninga z 16 sonarami; jednocześnie wystrzelują cztery sonary, po jednym z każdej ćwiartki pierścienia. Pomaga to w przesłuchaniu, ale nie jest to kompletne ani niezawodne rozwiązanie. Jeśli można zmienić częstotliwość dźwięku sonaru i szybkość strzelania (co na ogół nie ma miejsca), można zastosować wyrafinowane techniki aliasingu.

Jeden z badaczy, Monet Soldo, opowiada historię opracowania reaktywnego robota mobilnego dla firmy IBM T.J. Laboratoria Watson pod koniec lat 80-tych. Robot używał sonaru jako swoich głównych czujników, a Soldo miał zapisane zachowania, aby skutecznie prowadzić robota przez drzwi, pokoje i korytarze z rozsądną prędkością. Nadszedł dzień wielkiej demonstracji, która miała się odbyć nie na korytarzach laboratorium, ale w recepcji. Robot z powodzeniem wyszedł z laboratorium, wzdłuż korytarzy, a następnie wpadł w szaf, gdy dotarł do atrium. Soldo uruchomił się ponownie i spróbował ponownie, ale uzyskał ten sam wynik. Po wielu dniach prób debugowania kodu, zdała sobie sprawę, że to nie problem z kodem, to problem środowiskowy: większość recepcji w atrium składała się ze szklanych ścianek działowych. Odbicie lustrzane i przesłuchy powodowały halucynacje robota, choć za każdym razem w inny sposób. Miałem podobny problem z robotem szaleńczym, gdy mój robot oparty na sonarze zaczął nawigować w środowisku biurowym. W takim środowisku oczekiwano robota do poruszania się pomiędzy boksami biurowymi oddzielonymi przegrodami. Przegrody przykryto tkaniną, aby wytłumić odgłosy robotników. Niestety płótno pochłaniało również dźwięk z sonarów! Te historie podkreślają potrzebę rozważenia środowiska pracy czujnika i jego wpływu na sygnał. Wpływ problemów z odbiciami zwierciadlanymi i rozmowami krzyżowymi staje się łatwiejszy do zauważenia dzięki wykresom zwrotów sonaru nałożonym na obwód obszaru, w którym zostały one zrobione, patrz rysunek



Niektóre ściany są niewidoczne; inni są zbyt blisko. Odczyty czujnika są wysoce niepewne i w rezultacie są obecnie obsługiwane przy użyciu metod probabilistycznych, które zostaną omówione w rozdziale 14. Stożek 30° również stwarza problemy z rozdzielczością. Chociaż sonary często mają doskonałą rozdzielczość w głębi, mogą to osiągnąć tylko na duże odległości, jeśli obiekt jest wystarczająco duży, aby odesłać znaczną część fali dźwiękowej. Im dalej obiekt znajduje się od robota, tym większy musi być obiekt. Większość krzeseł biurowych i blatów nie ma prawie żadnej powierzchni dla czujnika, więc robot często nie zauważa ich obecności i wpada na nie. W praktyce do fałszywych odczytów sonaru prowadzi inny problem: moc. Wytworzenie fali dźwiękowej wymaga znacznego impulsu energii. Jeśli robot pracuje na niskich poziomach mocy, prawidłowy przebieg nie zostanie wygenerowany, a sygnał zwrotny będzie bezwartościowy. Jedną z metod eliminacji fałszywych odczytów, niezależnie od przyczyny, jest pobranie średniej z trzech odczytów (bieżący odczyt plus ostatnie dwa) z każdego czujnika. Ta metoda jest dość powszechna w przypadku robotów czysto reaktywnych, ale jest to proces ad hoc. Jak zobaczymy w dalszych rozdziałach, inne podejścia traktują odczytanie jako niepewne i stosują formalne techniki wnioskowania dowodowego w celu udoskonalenia odczytania. Te techniki niepewności są stosowane przez architektury działające zgodnie z paradygmatem hybrydowym

Jasne paski

Jasne paski, jasne paski lub detektory światła strukturalnego działają, wyświetlając kolorową linię (lub pasek), siatkę lub wzór kropek na otoczenie. Następnie zwykła kamera wizyjna obserwuje zniekształcenie wzoru na obrazie. Obecne kamery RGB-D są oparte na paskach świetlnych, ale zamiast widocznych linii systemy kamer wyświetlają losowe kropki światła podczerwonego, które nie są widoczne dla ludzkiego oka.

PRZESZKODA NEGATYWNA

Uproszczony przykład działania jasnego malowania pokazano na rysunku



a.



b.



c.

Na rysunku a paskowacz wyświetla cztery linie. Linie powinny występować w określonych, równomiernie rozmieszczonych rzędach na obrazie z kamery, jeśli powierzchnia jest płaska. Jeśli powierzchnia nie jest płaska, jak pokazano na rysunku b, linie będą miały przerwy lub nieciągłości. Algorytm wizyjny może szybko przeskanować każdy z wyznaczonych wierszy, aby sprawdzić, czy rzutowana linia jest ciągła, czy nie. Lokalizacja przerw w linii daje informację o wielkości przeszkody. Algorytm wizyjny może również patrzeć na lokalizację, w której pojawiają się przemieszczone segmenty linii, ponieważ odległość we współrzędnych obrazu jest proporcjonalna do głębokości obiektu. Względne położenie linii wskazuje, czy obiekt znajduje się nad płaszczyzną podłoża (przeszkoda), czy poniżej (dziura lub przeszkoda ujemna). Im więcej linii lub drobniejszej siatki, tym więcej informacji o głębokości. Z wielu powodów jasne paski są tańsze niż dalmierze laserowe. Po pierwsze, ponieważ wytwarzają linię lub wzór do zmierzenia, drogie detektory czasu wyłączenia są niepotrzebne. Wykrywanie odbywa się przez kamerę. Po drugie, wytworzenie grubej linii, którą może wykryć kamera, nie wymaga lasera. Zamiast tego można to zrobić za pomocą światła strukturalnego, metody optycznej, która pozwala „zwykłemu” światłu naśladować pożądane właściwości światła laserowego. Wreszcie jasne paski tworzą dość szorstki wzór; nie wyświetlają linii ani siatki na każdym pikselu obrazu. Oznacza to, że malowarka jest mniej wymagająca, a zatem tańsza w budowie. Jasne paski były popularne pod koniec lat 80. i na początku lat 90., a laboratoria, takie jak Laboratorium Robotyki Ogólnej, Automatyki, Wyczuwania i Percepcji Uniwersytetu Pensylwanii (GRASP Lab) pod kierownictwem Ruzeny Bajcsy, dały doskonałe wyniki w wyodrębnianiu informacji 3D. Jednak wysiłki te koncentrowały się na wykorzystaniu mapy głębi do rozpoznawania obiektu w warunkach laboratoryjnych. Wyniki nie przeniosły się szczególnie dobrze na roboty mobilne w mniej korzystnych warunkach. Roboty reaktywne nie zajmują się rozpoznawaniem, więc wiele algorytmów nie było szczególnie użytecznych lub nie było w stanie zapewnić szybkiej, odruchowej odpowiedzi. Ponadto w

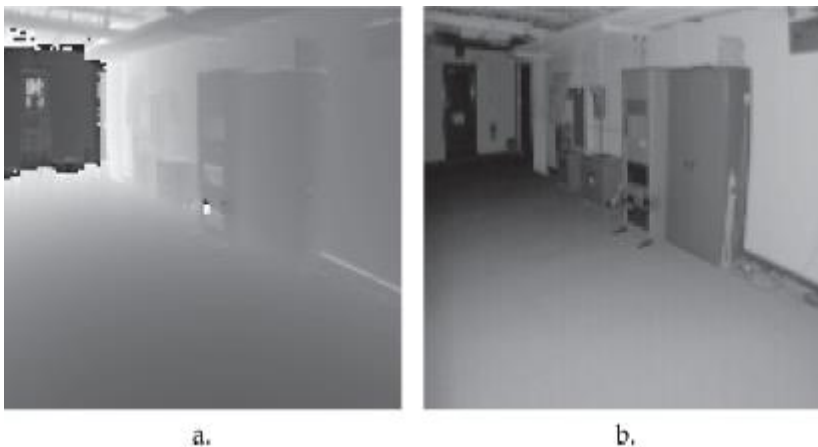
otwartym świecie obiekty były często tego samego koloru co emitowane światło lub wystarczająco blisko niego, aby zmylić system wizyjny. Ilość i kierunek oświetlenia mogą również zmylić malowarkę, ponieważ jasno oświetlone pomieszczenia utrudniają mu dostrzeżenie jasnego światła laserowego lub strukturalnego. Łazik Sojourner Mars wykorzystywał system świetlny do omijania przeszkód. System wyświetlał pięć linii przed pojazdem 198. Działało to dobrze, ponieważ Mars ma niewielkie zróżnicowanie kolorów i światła. Co ciekawe, urządzenie do pasowania światła używało jednego członka pary stereo, ale w przeciwieństwie do Marsokhoda, Sojourner nie używał stereo do nawigacji. Zamiast tego robot okresowo zabierał parę stereo do rekonstrukcji i tworzenia map na Ziemi.

LIDAR

Ultradźwięki wprowadzają energię akustyczną do otoczenia i mierzą czas przelotu sygnału do powrotu. Ta sama zasada może być używana z laserami; emitowana jest wiązka laserowa i mierzony jest współczynnik odbicia. W przeciwieństwie do wiązki sonaru, która ma bardzo szerokie pole widzenia (prawie 30°), laser wytwarza prawie nieskończenie małe pole widzenia. Jeśli laser jest skierowany w skanie, podobnie jak skan rastrowy na CRT, urządzenie może pokryć rozsądny obszar i wytworzyć obraz, gdzie funkcja obrazu generuje wartości głębi. Urządzenia wykorzystujące lasery do tworzenia mapy głębi lub obrazu są często nazywane radarem laserowym, ladarem lub lidarem. Mogą generować nawet milion pikseli zasięgu na sekundę, z zasięgiem 30 metrów i dokładnością kilku milimetrów. Mechaniczny komponent skanujący sprawia, że lidary są bardzo drogie, rzędu od 30 000 do 100 000 USD. Tańszym rozwiązaniem do nawigacji jest stworzenie planarnego dalmierza laserowego.

SEGMENTACJA ZAKRESU

Lidar tworzy dwa obrazy: intensywność i zasięg. Rysunek



przedstawia obrazy wykonane przez kamerę laserową Odetics (LADAR). Mapa intensywności jest zasadniczo czarno-białą fotografią, która mierzy intensywność światła odbitego lub pochłoniętego przez obiekty na scenie. Odpowiada to temu, jak ludzie postrzegają tę scenę. Funkcja obrazu dla obrazu zakresu reprezentuje głębię z kamery. Piksele, które są czarne lub mają wartość 0, są bliżej kamery zakresu niż białe piksele. Płaska podłoga zwykle pojawia się jako promieniujący zestaw półokręgów przechodzących od czerni (blisko) do bieli (daleko). Trygonometria jest następnie używana do obliczenia, że okręgi reprezentują płaską powierzchnię. Proces ten nazywa się segmentacją zakresu i może być dość trudno zrekonstruować powierzchnie na podstawie danych. W praktyce lidary mają pewne problemy. Rysunek pokazuje obszar na obrazie zakresu, który jest całkowicie czarny lub bardzo blisko. Ale jak widać na obrazie intensywności, obszar ten jest w rzeczywistości daleko. Podobnie czarna listwa pomiędzy ścianą a podłogą wydaje się być bardzo odległa na obrazie zakresu. Błędy wynikały z warunków poza zasięgiem, absorpcji światła (zwróconego za mało światła) lub z optycznego

odpowiednika odbicia zwierciadlanego (światło padające na rogi jest odbijane od odbiornika). Planarny dalmierz laserowy, emituje płaszczyznę światła laserowego. Zapewnia to cienki poziomy wycinek przestrzeni i tworzy zasadniczo wykres biegunowy o wysokiej rozdzielczości. Roboty początkowo były wyposażone w laser SICK montowany równolegle do podłogi. Było to przydatne do omijania przeszkód (o ile przeszkoda była wystarczająco wysoka, aby złamać płaszczyznę rzutowaną przez laser), ale nie była szczególnie pomocna przy wydobywaniu informacji 3D. Podobnie jak w przypadku sonarów, roboty narażone były na dekapitację przez przeszkody, takie jak stoły, które nie pojawiały się w polu widzenia czujnika zasięgu, ale które mogłyby uderzyć w uchwyt lub antenę czujnika. Aby rozwiązać ten problem, naukowcy zaczęli ostatnio montować planarne dalmierze laserowe pod niewielkim kątem do góry. Gdy robot porusza się do przodu, ma inny widok na nadchodzące przeszkody. W niektórych przypadkach naukowcy zamontowali dwa dalmierze laserowe, jeden lekko przechylony w górę, a drugi lekko w dół, aby zapewnić pokrycie nadwieszonych i negatywnych przeszkód.

KAMERA RGB-D

Kamery RGB-D to połączenie systemu pasów światła podczerwonego (z wyjątkiem pasów światła, które są wyświetlane losowo w kropki) ze zwykłą kamerą światła widzialnego RGB. Są one czasami nazywane kamerami RGB+D lub po prostu RGBD. Kamery RGB-D wyświetlają światło podczerwone odczytywane przez kamerę dostrojoną do częstotliwości światła podczerwonego. Kropki IR są niewidoczne dla ludzkiego oka. W ten sposób obok kamery na podczerwień można zamontować oddzielną kamerę RGB, aby zobaczyć tę samą scenę bez punktów i uczynić ją widoczną dla ludzkiego oka. Niedrogie kamery Microsoft Kinect i ASUS Xtion wykorzystują zasadniczo tę samą technologię w grach wideo i zostały natychmiast zaadoptowane do użytku przez robotyków. Jednak, jak zauważyli Suarez i Murphy,²⁰¹ kamery RGB-D nie są wystarczające dla ogólnej robotyki z trzech powodów. Po pierwsze, kamery RGB-D są zazwyczaj zoptymalizowane pod kątem rynku gier wideo, w którym kamera jest nieruchoma w pomieszczeniu, a gracz dostosowuje oświetlenie w pomieszczeniu i obraca system kamer, aż do uzyskania spójnych wyników. W robotach mobilnych robot może poruszać się po obszarach, w których panują znacznie inne warunki oświetleniowe niż w pomieszczeniu rekreacyjnym. Jasne światło i bezpośrednie światło słoneczne rozmywają projekcję IR. Po drugie, światło podczerwone jest aktywnym czujnikiem. W ten sposób obiekty o powierzchniach odbijających lub ostrych krawędziach i powierzchniach kątowych, które tworzą odbicie zwierciadlane, powodują, że obraz w podczerwieni jest zaszumiony. Po trzecie, kamery RGB-D są dostrojone pod kątem projekcji i odbioru w określonych zakresach. W przypadku sensora Kinect zasięg ten wynosi od 1,2 m do 3,5 m. Jeśli robot musi dostrzec głębokość bliżej niż 1,2 m lub dalej niż 3,5 m, sensor Kinect nie będzie przydatny.

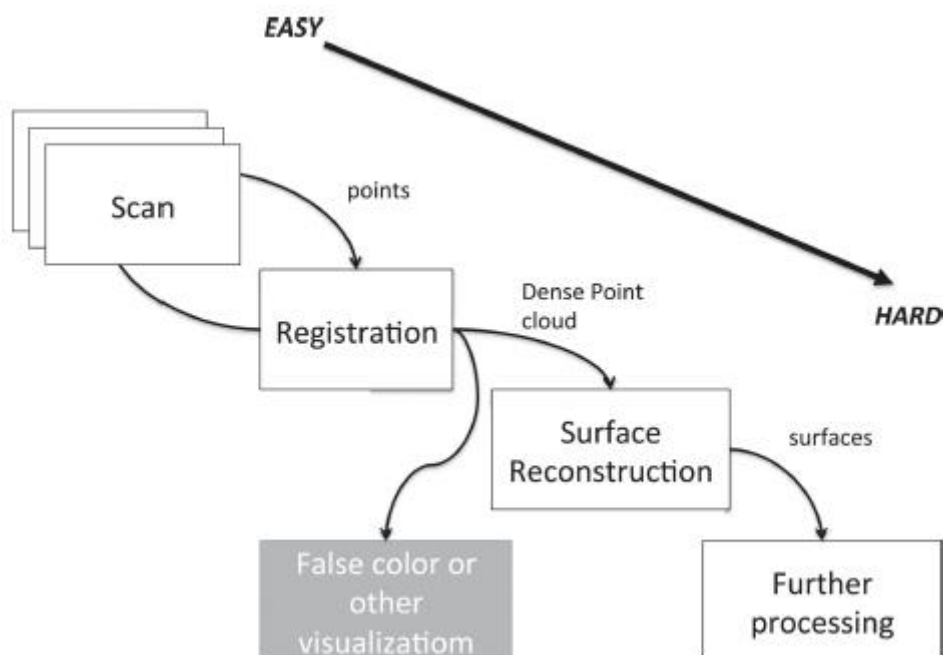
CHMURA PUNKTÓW

Chmury punktów są reprezentacją wiedzy o zasięgu do powierzchni. Chmura to zbiór punktów, zwykle tysięcy punktów. Chmura może być tablicą lub listą punktów. Każdy punkt ma albo współrzędną (x, y, z), zwaną wierzchołkiem, reprezentującą odległość między czujnikiem a powierzchnią obiektu lub wektor (kierunek, odległość). Sam punkt w rzeczywistości reprezentuje objętość przestrzeni, podobnie jak sygnał zwrotny sonaru, ale rozdzielczość jest na ogół o kilka rzędów wielkości wyższa niż ta, którą wytwarza czujnik ultradźwiękowy. Punkt może mieć również wartość, taką jak nasycenie światła, co może dać obraz podobny do czarno-białej fotografii. Chociaż chmura punktów technicznie odnosi się do zbierania punktów, termin ten przywołuje powiązane algorytmy i style wizualizacji do pracy z punktami. Chmury punktów są zwykle generowane przez czujniki aktywne, czujniki zasięgu lidarowego lub czujniki światła strukturalnego. Czujniki o aktywnym zasięgu, takie jak modele wyprodukowane przez SICK, Velodyne i Hokuyu emitują wiązkę laserową, a następnie rejestrują czas, w którym wiązka odbija się od powierzchni i powraca. Czujniki skanują albo w płaszczyźnie dwuwymiarowej, jak kartka

papieru, albo skanują w trzech wymiarach. Skanowanie w trzech wymiarach jest droższe, ponieważ jest znacznie bardziej skomplikowane optycznie i elektromechanicznie. Czujniki SICK zostały pierwotnie opracowane do pomiarów precyzyjnych, w których wystarcza 2D. Inne skanery 2D, czasami nazywane planarnymi dalmierzami laserowymi, takie jak Hokoyu, są ograniczone do 2D, aby obniżyć koszty. Czujniki światła strukturalnego zwykle emitują światło w paśmie częstotliwości bliskiej podczerwieni, które jest niewidoczne dla ludzkiego oka. Najbardziej znanym sensorem strukturalnego zasięgu światła jest Microsoft Kinect. Punkty mogą być generowane systematycznie, tak jak skan rastrowy, który tworzy tablicę tysięcy lub milionów punktów, lub czujnik może losowo wykrywać region w celu zmniejszenia liczby punktów na każdym skanie. Pierwotnie termin chmura punktów sugerował znacznie większą gęstość punktów niż te generowane z widzenia stereoskopowego — tak dużą, że do efektywnej pracy z dużą liczbą punktów potrzebna była inna klasa algorytmów. Niektóre algorytmy nawet nie próbują wykorzystać wszystkich punktów, ale zamiast tego losowo próbują punkty w chmurze, aby zmniejszyć czas potrzebny na obliczenia. Reprezentacja wiedzy dotycząca chmury punktów i powiązanych algorytmów do wydajnego przetwarzania punktów jest obecnie używana ogólnie. Piksele w obrazie głębi z geoprzestrzennych stereoskopowych obrazów satelitarnych lub metod fotogrametrycznych są obecnie uważane za chmury punktów, a nie macierze.

ZMYSŁ

Algorytmy pracy z chmurą punktów są często stosowane w procesie sekwencyjnym pokazanym na rysunku



Wysiłek obliczeniowy wzrasta, gdy proces przechodzi od zdobywania punktów do próby grupowania punktów w odpowiednie powierzchnie, które odtwarzają środowisko. Pierwszym krokiem w tym procesie jest wykrycie przez czujnik otoczenia i wygenerowanie zestawu punktów.

REJESTRACJA

Po drugie, ten zestaw punktów musi być zarejestrowany z innymi zestawami punktów. Na przykład pomieszczenie może wymagać wielu skanów, a te widoki muszą zostać połączone, aby utworzyć jeden

ujednolicony widok. Innym przykładem jest włączanie skanów w czasie, na przykład gdy robot porusza się w kierunku obiektu. Zwykle głównym elementem rejestracji jest przypisanie każdemu punktowi położenia x,y,z we współrzędnych bezwzględnych, ponieważ skan będzie raportowany we współrzędnych względnych lub egocentrycznych, które należy przekształcić we wspólny zestaw współrzędnych bezwzględnych. Rejestracja może również nałożyć obraz wizualny na chmurę punktów. Chmury punktów same w sobie są często tak gęste, że samo wyświetlanie punktów w różnych odległościach od robota fałszywym kolorem jest rozsądną wizualizacją głębokości. W przypadku pojedynczego obiektu lub rzadkiej sceny, punkty na powierzchni często znajdują się w tej samej odległości od punktu widzenia, a zatem mają ten sam kolor. Jednak w przypadku bardziej złożonych scen wizualizacja fałszywych kolorów wymaga od człowieka mentalnego podzielenia świata na powierzchnie.

REKONSTRUKCJA

Robot powieła wysiłek umysłowy w zamianie punktów na powierzchnie poprzez rekonstrukcję powierzchni. Zasadniczo punkty w chmurze są wierzchołkami płatów powierzchni. Jednym ze sposobów tworzenia powierzchni jest zastosowanie algorytmów siatki, takich jak triangulacja Delaunaya lub sześciiany marszowe, które próbują przekształcić wierzchołki w siatkę wielokątów lub trójkątów reprezentujących powierzchnie w chmurze punktów. Innym podejściem jest zastosowanie algorytmów niejednorodnych racjonalnych splajnów bazowych (NURBS), które próbują łączyć wierzchołki za pomocą krzywych. Po przekształceniu chmury punktów dla regionu w powierzchnie można zastosować dalsze przetwarzanie. Powierzchnie mogą stać się częścią rekonstrukcji 3D środowiska i służyć do rozpoznawania obiektów lub planowania ścieżki w środowisku. Inne przetwarzanie obejmuje wyodrębnianie pozy dłoni lub ramienia do kontrolowania gry (lub robota).

Studium przypadku: Hors d'Oeuvres, Anyone?

Zgłoszenie University of South Florida (USF) w 1999 roku w konkursie AAI Mobile Robot Competition Hors d'Oeuvres, Anyone? Wydarzenie zawiera studium przypadku wyboru czujników, konstruowania zachowań reaktywnych i wykorzystywania behawioralnej fuzji czujników. We wpisie wykorzystano dwa współpracujące roboty. Celem było rozszerzenie zakresu robotycznego wykrywania poprzez zastosowanie sześciu modalności wykrywania z 40 różnymi urządzeniami fizycznymi na jednym robocie i czterech modalności z 23 urządzeniami na drugim.

Krok 1: Opisz zadanie. „Przystawki, Ktoś?” Wydarzenie wymagało w pełni autonomicznych robotów, aby krążyły po recepcji na konferencji AAI z tacą z przekąskami, znajdowały i zbliżały się do ludzi, wchodziły z nimi w interakcję i uzupełniały tacę. Każdy robot był oceniany za pokrycie obszaru, zauważając, kiedy taca wymagała uzupełnienia, naturalnie wchodzić w interakcję z ludźmi, mieć wyraźną osobowość i rozpoznawać VIP-ów. W zgłoszeniu USF wykorzystano dwa roboty, Borg Shark i Puffer Fish, przebrane przez Departament Sztuki USF, aby przyciągnąć uwagę i zasugerować różne osobowości. Borg Shark był robotem-serwerem, który poruszał się po widowni, podążając wcześniej zaplanowaną trasą. Robot zatrzymywał się i podawał w regularnych odstępach czasu lub za każdym razem, gdy z tacy wyjmowano smakołyk. Używał syntezy mowy do nadawania plików audio, zachęcając widzów do wyjęcia smakołyku z ust, ale nie miał możliwości usłyszenia i zrozumienia ludzkich poleceń w języku naturalnym. Aby wchodzić w bardziej naturalną interakcję z ludźmi, Borg Shark próbował utrzymywać kontakt wzrokowy z ludźmi. Jeśli zobaczył osobę, oszacował położenie we współrzędnych obrazu, gdzie znajdowała się kolorowa plakietka VIP-a w stosunku do twarzy osoby. Kiedy Borg Shark prawie skończył się jedzeniem, dzwoniła przez radio Ethernet do swojego robota-pomocnika, Puffera Fisha. Ryba Rozdymkowa była nieruchoma w trybie uśpienia, wdychając i wydychając powietrze przez swoją nadmuchiwaną spódnicę i obracając aparaty, jakby unikając

tłoczących się ludzi. Kiedy Puffer Fish się obudziła, udała się z pełną tacą jedzenia (umieszczoną na jej stojaku przez człowieka) na współrzędne podane jej przez Borg Shark. Szukała również charakterystycznego niebieskiego kostiumu Borga Sharka, używając zarówno liczenia martwych, jak i wyszukiwania wizualnego, aby przejść do celu. W odległości dwóch metrów od Borg Shark, Puffer Fish zatrzymał się. Człowiek fizycznie zamieniłby się tacami, a następnie kopnął zderzaki, aby zasygnalizować, że transfer się skończył. Borg Shark wznowił swój cykl podawania, podczas gdy Puffer Fish wróci do swojej domowej stacji napełniania. Oba roboty miały omijać wszelkie przeszkody: stoły, krzesła, ludzi. Ponieważ ludzie mieli tendencję do otaczania robotów, uniemożliwiając pokrycie obszaru lub ponowne napełnianie, roboty miały różne reakcje. Borg Shark, który został zaprogramowany do bycia sprytnym, ogłosił, że nadchodzi i zacznie się poruszać. Rozdymkowata Ryba, o zrzedliwej, ponurej osobowości, głośno narzekała, głośno napompowując spódnicę, a następnie gwałtownie szarpnęła do przodu, zwykle powodując, że widzowie cofali się i dawali jej miejsce.

Krok 2: Opisz roboty. Roboty użyte do wejścia to bazy Nomad 200 z niestandardowymi czujnikami. Oryginalne czujniki Borg Shark obejmowały parę kolorowych kamer zamontowanych na głowicy pan-tilt, nadmiarowe pierścienie sonaru i laser planarny SICK,

Czujniki dla Borg Shark.

Typ : Modalność : Urządzenia

exteroceptive: wizja: 2 kamery

exteroceptive : laser : 1 planarny dalmierz

exteroceptive : sonar : 15 ultradźwięków (górny), 15 ultradźwięków (dolny)

exteroceptive: dotykowy: 1 przełącznik zderzaka;

proprioceptywne : enkodery silnika : 3 napęd, sterowanie, sterowanie wieżyczką , 2 pan, sterowanie pochyleniem

Czujniki dla Puffer Fish.

Typ : Modalność : Urządzenia

exteroceptive: wizja: 2 kamery

exteroceptive: sonar: 15 ultradźwięków

exteroceptive: dotykowy: 1 przełącznik zderzaka;

proprioceptywne : enkodery silnika : 3 napęd, sterowanie, sterowanie wieżyczką , 2 pan, sterowanie pochyleniem

Krok 3: Opisz środowisko. Otoczenie było areną kongresowego ze zmiennym oświetleniem i wieloma ludźmi ubranymi swobodnie. Głównym wyzwaniem percepcyjnym robotów było utrzymywanie kontaktu wzrokowego z ludźmi i ustalenie, czy dana osoba jest VIP-em. Odznaki VIP-ów były oznaczone charakterystycznymi kolorowymi wstążkami, więc jeśli robot był pewien, że patrzy we właściwe miejsce, kolor dawałby VIP-owi. Jednak kolory wstążek były dość powszechne. Gdyby robot szukał wstążek mniej więcej na wysokości klatki piersiowej, prawdopodobnie znalazłby w tłumie koszulkę w tym kolorze, ubraną w jaskrawe stroje turystyczne (konkurencja odbywała się w Orlando). Spowodowałyby to błędną identyfikację VIP, tracąc punkty. Innym podejściem było użycie sonarów do znalezienia osoby poprzez czytanie

zasięg, kierując kamerę w tym kierunku i nachylając się pod kątem, pod którym będą znajdować się oczy osoby o średnim wzroście. Ten mechanizm skupienia uwagi może być również wykorzystany do sprawdzenia prawdopodobnej lokalizacji odznaki, jeśli jest to pożądane. Jednak z doświadczenia wiadomo było, że ludzie rzadko stoją wzdłuż akustycznej osi sonaru. Jeśli jedna osoba zostałaby podniesiona przez dwa sonary, robot szukałby przez 10 sekund w lewo od osoby, a następnie 10 sekund w prawo. Gdyby obecnych było wiele osób, robot wydawałby się jeszcze bardziej dysfunkcyjny, a mechanizm skupienia uwagi nie działałby poprawnie. Lepszym rozwiązaniem byłoby dla robotów wykrywanie osoby za pomocą wzroku. Zauważ, że wykrywanie to nie to samo, co rozpoznawanie. Wykrywanie oznacza, że robot jest w stanie zidentyfikować twarz, która reaguje. Rozpoznanie oznacza etykietowanie twarzy i możliwość jej późniejszego rozpoznania, co jest funkcją deliberatywną. Czy istnieje prosta afordancja wizualna twarzy? W rzeczywistości ludzka skóra, niezależnie od pochodzenia etnicznego, jest niezwykle podobna kolorystycznie do systemu wizyjnego. Gdy robot znajdzie kolorowy obszar o wielkości i kształcie głowy, będzie mógł bardziej niezawodnie znaleźć odznaki VIP. Inną szansą na afordancję była nawigacja Puffer Fish do Borg Shark. Chociaż rozdymkowane ryby otrzymałyby współrzędne Borg Shark, jest mało prawdopodobne, aby rozdymkowane mogły niezawodnie nawigować do Borg Shark, używając tylko martwego liczenia. Współrzędne prawdopodobnie były nieprawidłowe z powodu własnego dryfu Borg Shark w czasie. Wtedy Puffer Fish nagromadziłby martwy błąd w rachunkach, tym bardziej, gdyby musiał się zatrzymać, zacząć i unikać ludzi. Dlatego zdecydowano, że Ryba Rozdymkowa powinna szukać rekina Borga. Nakrycie głowy Borga Sharka zostało celowo duże i miało charakterystyczny niebieski kolor, aby zapewnić widoczność nad tłumem i zmniejszyć prawdopodobieństwo fiksacji rozdymki na czyjejs koszuli.

Krok 4-7: Projektuj, testuj i udoskonalaj zachowania. Na wybór czujników do innych zachowań, takich jak usuwanie przysmaków, miała wpływ fizyczna lokalizacja czujników. Na platformie badawczej zamontowano laser SICK dla Borg Shark. Platforma badawcza, nominalnie górna część robota, znajdowała się na wysokości dłoni, co czyni ją logicznym miejscem do zamocowania tacy do przechowywania żywności. Było oczywiste, że laser może być użyty do monitorowania obszaru tacy na żywność. Inne zespoły próbowały różnych metod, takich jak posiadanie kolorowej tacy i liczenie widocznego obszaru tego koloru (więcej koloru oznacza mniej smakołyków na tacy). Innym podejściem było zbudowanie wagi i monitorowanie zmian wagi tacki. Kostiumy robotów wpłynęły pośrednio na zestaw czujników. W ramach nadania robotom osobowości każdy robot miał kostium. Ryba Rozdymkowa miała nadmuchiwaną spódnicę, która wydymała się, gdy robot był zatłoczony lub zdenerwowany. Zespół musiał przetestować empirycznie i zmodyfikować spódnicę, aby upewnić się, że nie będzie kolidować z odczytami sonaru. Początkowe zachowania dla Borg Shark podano w tabeli

| Releaser | Behavior | Motor Schema | Percept | Perceptual Schema |
|---|--------------|---|-----------------------------------|---|
| always on | avoid() | vfh() | most-open-direction | polar-plot(sonar) |
| FOOD-REMOVED= treat-removal(laser) | track-face | center-face(face-centroid) track-face() check-VIP() | face-centroid ribbon-color | find-face(vision) look-for-ribbon(VIP-color) |
| SERVING-TIME-OUT, TRAY-FULL=bumper() | move-to-goal | pfields.attraction(waypoint) | waypoint | list of waypoints |
| FOOD-DEPLETED= treat-removal(laser) | track-face | center-face(face-centroid) | face-centroid | find-face(vision) |

a Puffer Fish w tabeli poniżej.

| Releaser | Behavior | Motor Schema | Percept | Perceptual Schema |
|--------------------------------------|----------------|-------------------------------------|---------------------|-------------------------|
| always on | avoid() | vfh() | most-open-direction | polar-plot(sonar) |
| AT-HOME= dead-reckoning(encoders) | sleep() | turn-camera-head() cycle-skirt() | obstacle | polar-plot(sonar) |
| AWAKE=radio-signal() | move-to-goal() | pfields.attraction(location) | relative-location | read-encoders() |
| AWAKE=radio-signal() | move-to-goal() | pfields.attraction(shark) | shark | find-shark-blue(camera) |
| TRAY-FULL=bumper() | move-to-goal() | pfields.attraction(home) | relative-location | read-encoders() |

Jak widać w tabelach, jedynym zachowaniem wykorzystującym jakąkolwiek formę fuzji czujników było przejście do celu w Puffer Fish, który miał dwie konkurujące ze sobą instancje celu powodujące rozszczepienie czujnika. Zachowanie vfh to zachowanie polegające na unikaniu przeszkód przy użyciu wykresów biegunowych pochodzących z modeli opisanych w rozdziale 15. Gdy zespół testował zachowania indywidualnie, zidentyfikowano problemy związane z zachowaniami typu znajdź twarz i usuwanie leczenia. Chociaż kolor był rozsądną afordancją dla twarzy, algorytm często zwracał fałszywe negatywy, brakujące twarze, chyba że znajdowały się w jasnym świetle. W międzyczasie laser czasami wydawał się odbijać od zębów kostiumu, generując również fałszywie dodatnie wyniki, i w ponad 75% przypadków laser chybił dłoni osoby, jeśli ruch ręki był szybki. Stawki są pokazane poniżej:

Czujnik logiczny: Fałszywe pozytywy: Fałszywe negatywy

Znajdowanie twarzy : 1,7% : 27,5%

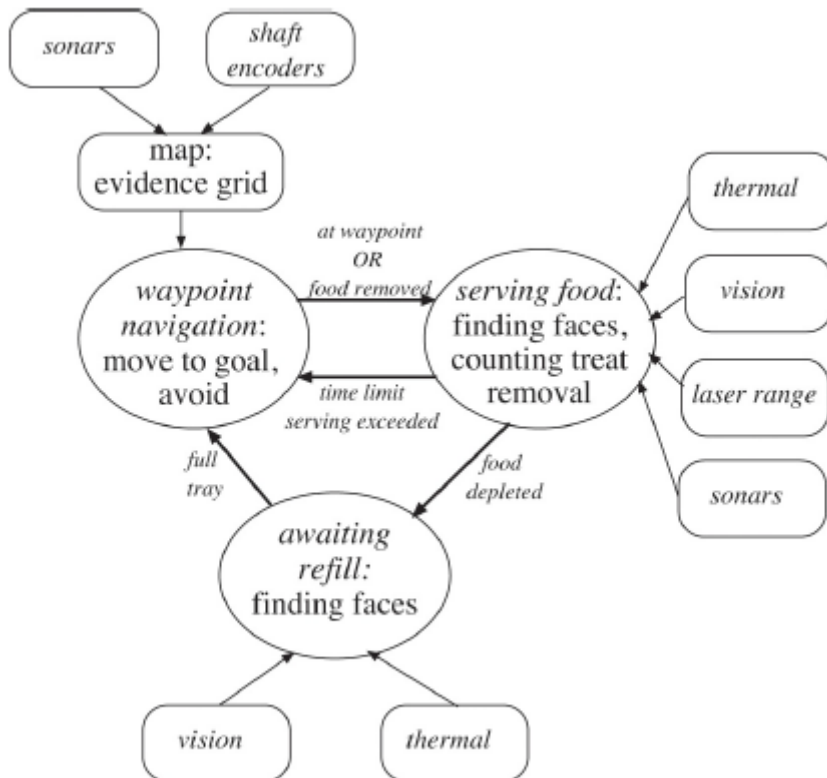
Licznik żywności: 6,7%: 76,7%

Rozwiązaniem problemu znalezienia twarzy było wykorzystanie innej afordancji człowieka, używanej przez komary: ciepła. Problemem była częściowa segmentacja człowieka, ponieważ regiony kandydujące były odrzucane ze względu na zbyt małe. Ciepło byłoby dobrym kryterium decyzji. Jeśli region kandydujący był kolokowany z regionem gorącym, był ogłaszany jako twarz. Na szczęście zespołowi udało się przenieść termometr cyfrowy E2T używany na innym robocie do Borg Shark. Czujnik termiczny pokazany na rysunku 11.17 był przeznaczony do określania temperatury osoby w kontakcie, ale był w stanie wykryć wzrost temperatury powyżej temperatury otoczenia od osoby znajdującej się w odległości do dwóch metrów. Dlaczego system wizyjny nie został po prostu zastąpiony czujnikiem termicznym? Ta sytuacja nawiązuje do atrybutów czujników i ich dopasowania do otoczenia. Czujnik termiczny ma pole widzenia 2° i opóźnienie pomiaru, przez co jest zbyt wolny, aby szybko skanować pomieszczenie w poszukiwaniu ludzi. Zamiast tego system wizyjny obejmował znacznie szersze pole widzenia i mógł wygenerować niewielką listę regionów kandydujących. Następnie, gdy kamera obracała się w kierunku największego obszaru, odczyty sondy termicznej mogły określić, czy to naprawdę osoba, czy nie. Problem z szacowaniem ilości żywności został znacznie zredukowany przez prostą funkcję AND z sonarami. System liczył smakotyki jako usunięty tylko wtedy, gdy przed tacą był odczyt z bliskiej odległości w tym samym czasie, gdy laser wykrył ruch ręki. Wskaźnik fałszywych odczytów znacznie spadł, jak widać poniżej:

| Logical sensor | W/O Fusion | | Fusion | |
|----------------|------------|-------|--------|------|
| | FP | FN | FP | FN |
| Face-Find | 1.7% | 27.5% | 2.5% | 0% |
| Food-Count | 6.7% | 76.7% | 6.7% | 1.7% |

W tym momencie warto zrobić krok w tył i przyjrzeć się wyczuwaniu Hors d'Oeuvres, Ktoś? wpis w zakresie atrybutów wymienionych w rozdziale 10. Przypomnijmy, że atrybutami oceny przydatności pojedynczego czujnika były pole widzenia i zasięg, dokładność, powtarzalność i rozdzielczość, czas reakcji w domenie docelowej, zużycie energii, niezawodność i rozmiar. Problemem było pole widzenia

i zasięg czujników, co widać po różnicach w znajdowaniu twarzy za pomocą wizji i czujników termicznych. Kamera miała znacznie lepsze pole widzenia niż czujnik termiczny, więc kamera została wykorzystana do skupienia uwagi czujnika ciepła. Powtarzalność była wyraźnie problemem dla lasera z jego wysokim wskaźnikiem wyników fałszywie dodatnich/fałszywie ujemnych. Sonar nie można było wykorzystać do szacowania położenia twarzy, ponieważ rozdzielczość była zbyt gruba. Każdy z czujników miał rozsądną reakcję z punktu widzenia sprzętu, chociaż algorytmy mogły nie być w stanie ich wykorzystać. Zużycie energii nie stanowiło problemu, ponieważ wszystkie czujniki były włączone przez cały czas ze względu na sposób budowy robotów. Niezawodność i rozmiar sprzętu nie były poważnymi względami, ponieważ sprzęt był już w robotach.



Algorytmiczne wpływy na konstrukcję czujnika obejmowały złożoność obliczeniową i niezawodność. Oba były zdecydowanie czynnikiem w projektowaniu schematów percepcyjnych dla zachowań reaktywnych. Roboty posiadały sprzęt do obsługi zakresu stereo (dwie kamery z dedykowanymi framegrabberami). Zakres stereo może być używany do znajdowania twarzy, ale wysoka złożoność obliczeniowa, nawet z ówczesnym procesorem klasy Pentium, nie mogła przetworzyć algorytmu w czasie rzeczywistym. Problemem była również niezawodność. Algorytm rozpoznawania twarzy był bardzo zawodny, nie z powodu aparatu, ale dlatego, że algorytm nie był dobrze dostosowany do otoczenia i w rezultacie wykrywał obce plamy. Wreszcie, zestaw czujników można ogólnie ocenić pod względem prostoty, modułowości i nadmiarowości. Zestawy czujników dla obu robotów można uznać za proste i modułowe, ponieważ zawierały kilka oddzielnych czujników, w większości dostępnych na rynku, które mogły działać niezależnie od siebie. Zestaw czujników wykazywał wysoki stopień fizycznej redundancji: jeden robot miał podwójne pierścienie sonaru, sonary, laser i para kamer mogły być używane do określania odległości, ignorując rozmieszczenie zębów rekina i złożoność obliczeniową stereo algorytmu zakresu. Istniała również duża ilość nadmiarowości logicznej, która została wykorzystana poprzez wykorzystanie behawioralnej fuzji czujników.

Podsumowanie

Zbadaliśmy Jaka jest różnica między głębokością a zasięgiem? i Dlaczego robotom tak trudno jest wyczuć głębokość i zasięg? Nawigacja i manipulacja wymagają eksterocepcji, dzięki której robot obserwuje otoczenie. W szczególności robot generalnie potrzebuje percepcji głębi na dużych odległościach do nawigacji; jest to ogólnie określane jako wykrywanie zasięgu. Preferowani są laserowcy lub Lidar, ale mają one wady, takie jak koszt, rozmiar i waga. Dają one wysoką gęstość punktów $\{x, y, z, d\}$, co prowadzi do reprezentacji chmury punktów i powiązanych algorytmów przetwarzania do rejestracji i rekonstrukcji. Kamery RGBD, takie jak te zapoczątkowane przez Microsoft Kinect, są tanie, lekkie i małe, ale opierają się na czujniku podczerwieni, który został starannie zaprojektowany do wewnętrznych pokoi gier i opierają się na konkretnych algorytmach stworzonych dla osób stojących w określonym miejscu. odległość od czujnika. Kamery RGB-D nie są czujnikami ogólnego przeznaczenia i mają poważne problemy z oświetleniem zewnętrznym oraz powierzchniami lustrzanymi i odbijającymi światło. Zbadano ultradźwięki i widzenie komputerowe, przy czym ultradźwięki zostały w dużej mierze odrzucone z powodu niskiej rozdzielczości i zawodności. Jednak reprezentacja siatki zajętości i wykorzystanie rozumowania dowodowego do radzenia sobie z niepewnością nadal występują w wykrywaniu. Widzenie komputerowe oferuje wiele podejść do wydobywania głębokości i wykrywania powierzchni, w tym używania par stereo i obliczania głębokości na podstawie X. Głębokość jest zazwyczaj na bliskie odległości do manipulacji. Czujniki i algorytmy zasięgu i głębokości mają wady, takie jak koszt obliczeniowy, wydatki finansowe, niezdolność do pracy w różnych warunkach oświetleniowych oraz brak algorytmów do solidnego przekształcania nieprzetworzonych danych z czujnika w dokładne reprezentacje 3D.