

Rozpoznawanie obiektów trójwymiarowych na podstawie danych RGBD wykorzystując podejście Bag of Words

Adam Kosiorek

September 22, 2013

Abstract

Abstrakt

1 Wstęp

1.1 Motywacja

Rozwój nowych technologii pozwala na gromadzenie oraz przetwarzanie bardzo dużych ilości danych. Aparaty i kamery cyfrowe przyczyniają się do powstawania milionów gigabajtów informacji każdego dnia. Co więcej, dobrej jakości urządzenia służące do obrazowania trójwymiarowego osiągnęły tak niskie ceny, że można je znaleźć w wielu gospodarstwach domowych. Jednym z nich jest *Microsoft Kinect* – korzystający z technologii *PrimeSense* – oparty o światło strukturalne sensor rejestrujący dane RGBD.

Nowe technologie przyczyniają się też do spadku cen robotów mobilnych. Można sobie wyobrazić, że w niedalekiej przyszłości zrobotyzowani asystenci zawitają w domach. Roboty takie będą musiały spełniać szereg wymagań związanych z wymogami bezpieczeństwa oraz wygodą użytkowania. W szczególności będą musiały być wyposażone w mechanizmy pozwalające na bezpieczną, a zarazem efektywną interakcję z otoczeniem — w tym z ludźmi.

Zapewnienie poprawnej interakcji z otoczeniem wymaga spełnienia wielu warunków. Między innymi są to:

- Interfejs człowiek-maszyna powinien pozwalać na swobodną komunikację z robotem.
- Posiadanie wiedzy na temat aktualnego położenia oraz dysponowanie mapą otoczenia - problem opisywany w literaturze jako *SLAM* — *Simultaneous Localization And Mapping*
- Zdolność semantycznej klasyfikacji miejsca, w którym robot się znajduje
- Zdolność semantycznej klasyfikacji pojedynczych obiektów

W przypadku semantycznej klasyfikacji otoczenia robota można zauważyć, że rejestracja elementów otoczenia wykorzystując np. skaner laserowy

będzie niemożliwa. Skanery takie mają ograniczony zasięg, a w przypadku przebywania w otwartej przestrzeni większość obiektów może znaleźć się poza zasięgiem skanera. Prowadzi to do utrudnienia bądź niemożności wykorzystania informacji przestrzennej do rozpoznawania otoczenia robota. W przypadku semantycznej klasyfikacji obiektów sytuacja przedstawia się inaczej — sam fakt wystąpienia problemu dowodzi, że obiekt znalazł się w zasięgu sensorów. Możliwa jest więc rejestracja chmury punktów opisująca badany przedmiot oraz wykorzystanie informacji przestrzennej do klasyfikacji tegoż.

1.2 Cel pracy

Celem niniejszej pracy jest napisanie aplikacji służącej do semantycznej kategoryzacji obiektów trójwymiarowych. Kategoryzacja powinna odbywać się w oparciu o dane RGBD oraz wykorzystywać reprezentację Bag of Words. Aplikacja powinna działać w czasie rzeczywistym, a jej skuteczność powinna pozwalać na wykorzystanie jej w rzeczywistych robotach mobilnych.

1.3 Zakres pracy

W pracy zostały przyjęte następujące założenia projektowe:

- System operuje na danych RGBD
- Analizowane obrazy powinny zostać przygotowane w ten sposób, że rozpoznawany obiekt powinien wypełniać ponad połowę powierzchni zdjęcia
- Kategoryzacja obiektów odbywa się z pominięciem segmentacji obrazu.

Do realizacji projektu wykorzystano biblioteki: *OpenCV*, *PointCloudLibrary*, *Boost*. Wszystkie wykorzystywane algorytmy przetwarzania obrazu oraz uczenia maszynowego zostały zaimplementowane przez osoby trzecie, a w większości pochodzą z wymienionych bibliotek.

Zakłada się, że zostaną wykorzystane implementacje algorytmów uwzględniające wielowątkowość, w celu przyspieszenia działania programu. Ponadto możliwe jest wykorzystanie platformy CUDA w celu

porównania wydajności i dalszego przyspieszenia obliczeń.

2 Przegląd literatury

3 Proponowane podejście

W poniższej pracy wykorzystuje się podejście Bag of Words (BoW) połączone z algorytmami uczenia nadzorowanego. Model BoW ma tę zaletę w stosunku do surowych zdjęć lub chmur punktów, iż zmniejsza tzw. semantic gap - lukę znaczeniową pomiędzy zdjęciem lub niskopoziomowymi cechami charakteryzującymi obraz a wysokopoziomowe koncepcjami, nadającymi zdjęciu znaczenie semantyczne np. zachód słońca, człowiek.

4 Opis zastosowanych rozwiązań

5 Eksperymenty

6 Podsumowanie

References

name last name title title et al journal volume ISBN ISSN url numeral