

Politechnika Warszawska
Wydział Mechatroniki

Rozpoznawanie obiektów 3D na podstawie danych RGBD

Adam Kosiorek
pod kierunkiem prof. dr hab. B. Siemiątkowskiej

11 lutego 2014

Spis treści

Założenia i zakres pracy

Podejście Bag of Words

Projekt Aplikacji

Bazy danych

Wyniki i podsumowanie

Założenia

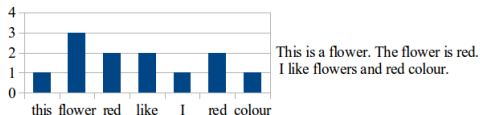
Założeniem pracy jest zbadanie zagadnienia klasyfikacji metodą Bag of Words obiektów 3D na podstawie zdjęć RGBD. Zdjęcia powinny być reprezentowane w formie chmur punktów. Ponadto zakłada się, że:

- ▶ Analizowane dane pochodzą z kamery RGBD Microsoft Kinect
- ▶ Implementacja algorytmu w języku C++ z wykorzystaniem bibliotek OpenCV i/lub PointCloud Library
- ▶ Testy aplikacji na ogólnie dostępnej, naukowej bazie danych

Zakres pracy

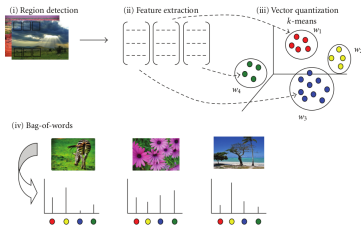
1. Przegląd istniejących rozwiązań dotyczących klasyfikacji obiektów na podstawie chmur punktów
2. Opracowanie algorytmu klasyfikacji obiektów na podstawie chmur punktów
3. Implementacja algorytmu w języku C++
4. Przeprowadzenie testów szybkości oraz skuteczności klasyfikacji
5. Opracowanie wniosków końcowych

Bag of Words - Wprowadzenie



- ▶ Częstość występowania słów
- ▶ Usunięcie gramatyki, kolejności słów
- ▶ Histogram jako forma pośrednia
- ▶ Wymaga utworzenia słownika
- ▶ Reprezentacja rzadka w przypadku dużego słownika
- ▶ Używany m. in. do znajdowania rozkładu tematów w obrazie

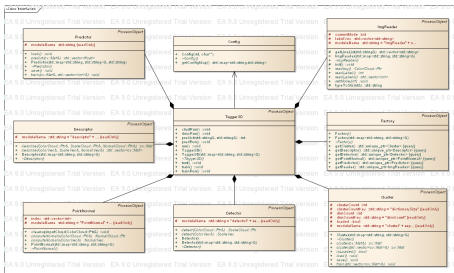
Bag of Words w obrazach



Grafika pochodzi z C. Tsai.
*Bag-of-words representation in
 image annotation: A review.*
ISN Artificial Intelligence, 2012.

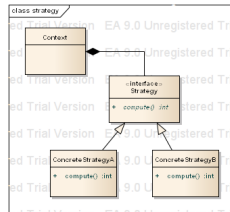
1. Wykrycie punktów charakterystycznych
 - ▶ współrzędne (x, y) oraz skala obrazka
2. Opisanie otoczenia wykrytych punktów
 - ▶ opis jednoznaczny, zazwyczaj poprzez przestrzenny histogram gradientów
3. Budowanie słownika i kwantyzacja
 - ▶ punkty układają się w obszary, które można wykryć - klasteryzacja
4. Klasyfikacja
 - ▶ klasyfikator uczący się - Support Vector Machine, Modele graficzne, Boosting

Projekt Aplikacji



Architektura

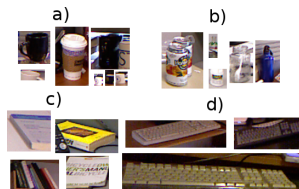
- ▶ Główny obiekt *Tagger3D* zarządza całą aplikacją
- ▶ Fabryka odpowiedzialna za tworzenie innych obiektów, konfigurowana w czasie działania programu
- ▶ Każdy etap procesu BoW ma wydzielony interfejs - wzorec projektowy *Strategia*
 - ▶ Detector, Descriptor itd.



Strategia

- ▶ Strategia umożliwia wymianę algorytmów w czasie działania programu, w sposób niewidzialny dla aplikacji
- ▶ Konfiguracja w pliku tekstowym
- ▶ Obiekt *IoUtils* oparty o szablonny obsługuje operacje wejścia/wyjścia

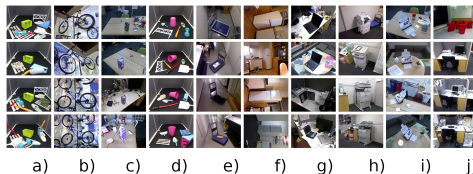
Berkely 3D Object Dataset



Rysunek: Obiekty z B3DO. Zdjęcia w oryginalnej jakości. a) kubek b) butelka c) książka d) klawiatura

- ▶ 78 kategorii obiektów
- ▶ Wiele obiektów na jednym zdjęciu
- ▶ Niska jakość zdjęć, wiele zdjęć niedoświetlonych, zaszumionych
- ▶ Obiekty przysłonięte
- ▶ Kolorowe zdjęcia i mapy głębi; Adnotacje w XML
- ▶ Skrypty w Pythonie i C++ do ekstrakcji pojedynczych obiektów i podziału na kategorie.
- ▶ Od 1 do 299 instancji w jednej kategorii
- ▶ Losowy wybór 8 kategorii tak, żeby było > 50 instancji w każdej kategorii

University of Tokyo Dataset



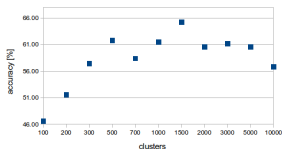
Rysunek: The Tokyo dataset: a) kosz b) rower c) pudełko d) wiadro e) wózek f) lodówka g) notebook h) drukarka i) skaner j) scena

- ▶ 10 kategorii
- ▶ 1 obiekt (oznaczony) na zdjęciu
- ▶ Mogą być obiekty nieoznaczone, których nie należy brać pod uwagę
- ▶ Zdjęcia dobrej jakości, dobrze naświetlone
- ▶ Wszystkie zdjęcia w kategorii to instancje jednego obiektu
- ▶ Skomplikowane sceny
- ▶ Małe różnice pomiędzy klasami obiektów
 - ▶ skaner i drukarka
 - ▶ koszyk i wiadro
- ▶ Dane w formie kolorowe: zdjęcie + plik .csv z absolutnymi współrzędnymi wszystkich pixeli
- ▶ Skrypty Python i C++ - przekształcenie do chmury punktów w formacie PCD

Porównanie algorytmów

Descriptor	Accuracy [%]		
	ISS3D	SIFT	Uniform Sampling
FPFH	65.22	56.52	56.07
PFH	59.32	54.83	59.50
PFHRGB	63.35	52.34	53.27

Porównanie algorytmów.



Wpływ rozmiaru słownika na skuteczność klasyfikacji.

- ▶ Wstępna selekcja algorytmów na podstawie przeglądu literatury
- ▶ Rodzina deskryptorów PFH uzyskuje najlepsze wyniki w dopasowywaniu chmur punktów:

PFH Algorytm bazowy, 125 wymiarów

FPFH Algorytm przybliżony, 33 wymiary

PFHRGB Algorytm uwzględnia kolor, 250 wymiarów

- ▶ Detektory:

SIFT Najlepszy algorytm 2D

ISS Najlepsze wyniki w dopasowywaniu chmur punktów

US Najprostszy, równomierna siatka punktów

- ▶ Optymalny rozmiar słownika

Wyniki B3DO

Assigned to category Category	cup	book	sofa	chair	table	bottle	monitor	keyboard	Entries per class	Accuracy [%]
cup	59	0	0	0	0	5	0	0	64	92.19
book	17	14	0	0	0	7	0	2	40	35.00
sofa	0	0	17	3	1	0	0	0	21	80.95
chair	0	0	2	14	4	0	2	0	22	63.64
table	0	0	0	0	31	0	0	0	31	100.00
bottle	34	0	0	0	0	30	2	0	66	45.45
monitor	7	1	0	2	0	10	30	1	51	58.82
keyboard	8	4	0	0	0	0	0	15	27	55.56

Wynik: 65.22% Wyniki na bazie B3DO, detektor ISS, ekstraktor FPFH, słownik o wielkości 1500 słów.

- ▶ Im lepiej rozdzielone kategorie obiektów, tym wyższe wyniki
- ▶ Najwyższy wynik: stół, 100%
- ▶ Niskie wyniki wynikają bezpośrednio z podobieństwa między klasami:
 - ▶ książka, 35%, mylona z filiżanką i butelką
 - ▶ butelka, 45%, mylona z filiżanką

Wyniki Tokyo

Assigned to category Category \	box	cart	notebook	scene	scanner	basket	bucket	freezer	bicycle	printer	Entries per class	Averages [%]
box	11	0	2	0	0	0	0	0	0	0	13	84.62
cart	0	0	1	0	0	0	3	0	0	0	4	0.00
notebook	6	0	3	1	0	0	1	0	1	0	12	25.00
scene	0	2	2	4	0	0	1	0	1	0	10	40.00
scanner	2	0	0	0	0	0	0	0	0	0	2	0.00
basket	1	0	0	1	0	2	7	1	1	0	13	15.38
bucket	3	0	0	0	0	0	22	1	0	0	26	84.62
freezer	0	0	2	0	0	0	1	1	0	1	5	20.00
bicycle	1	0	0	0	0	0	1	0	33	0	35	94.29
printer	0	0	0	0	0	0	1	0	1	0	2	0.00
Average												62.30

Wynik: 62.30% Wyniki na bazie Tokyo, detektor ISS, ekstraktor PFH, słownik o wielkości 3000 słów.

- ▶ Wyraźna zależność między ilością instancji w danej klasie i skutecznością klasyfikacji
- ▶ Najwyższy wynik: rower, 94%, największa ilość obiektów
- ▶ Najniższe wyniki przy klasach, w których są odpowiednio 4, 2 i 2 obiekty:
 - ▶ wózek, 0%
 - ▶ skaner, 0%
 - ▶ drukarka, 0%
- ▶ Notebook jest mylony z pudełkiem
- ▶ Wiadro jest mylone z pudełkiem
- ▶ Kosz jest mylony z wiadrem

Podsumowanie

- ▶ Uzyskane wyniki na bazach B3DO i TOKYO są odpowiednio 5.21 i 6.23 razy lepsze od kategoryzacji losowej (odpowiednio 12.5% i 10%)
- ▶ Zrealizowano wszystkie założenia projektowe
- ▶ Wystąpiły następujące problemy
 - ▶ W zasadzie brak baz danych nadających się do zadania klasyfikacji obiektów na podstawie danych RGBD
 - ▶ Bardzo wymagające przetwarzanie chmur punktów ze względu na brak jakiejkolwiek struktury
 - ▶ Najlepsze algorytmy dostępne są tylko do 2D, brak możliwości ich prostego uogólnienia do 3D
- ▶ Możliwości poprawy skuteczności i/lub szybkości działania aplikacji:
 - ▶ Przeniesienie obliczeń do domeny 2D; RGB i D jako niezależne obrazy + złożenie wyników klasyfikacji
 - ▶ Trening klasyfikatorów na znacznie większych bazach danych
 - ▶ Przeniesienie obliczeń na GPU