

Laboratorium CAD/CAM

Rozpoznawanie obiektów 3D przez sensor Kinect

Przygotowali

Michał Cwienczek

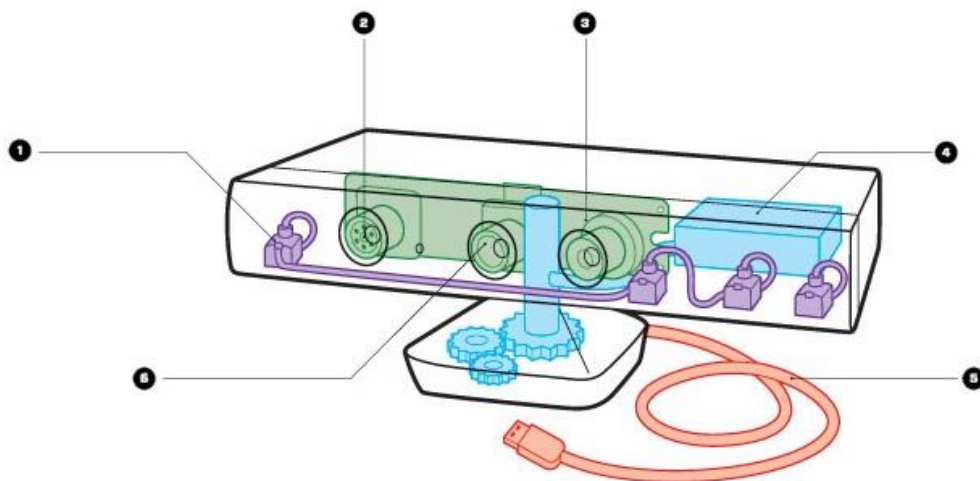
Michał Szewczak

1. Opis zadania

Zadanie polegało na przetworzeniu obrazu z kamery/kamer sensora kinect w taki sposób, aby rozpoznać obiekty znajdujące się przed sensorem. Do tego celu użyliśmy obrazu z kamer głębokości (chmura punktów głębokości) oraz obrazu z kamery zwykłej.

2. Sensor Kinect

Kinect jest urządzeniem stworzonym przez firmę Microsoft na potrzeby konsoli Xbox 360. Dostarcza on informacji o topologii pomieszczenia, w którym się znajduje, informacji o szkieletach widocznych graczy oraz rozpoznaje mowę.



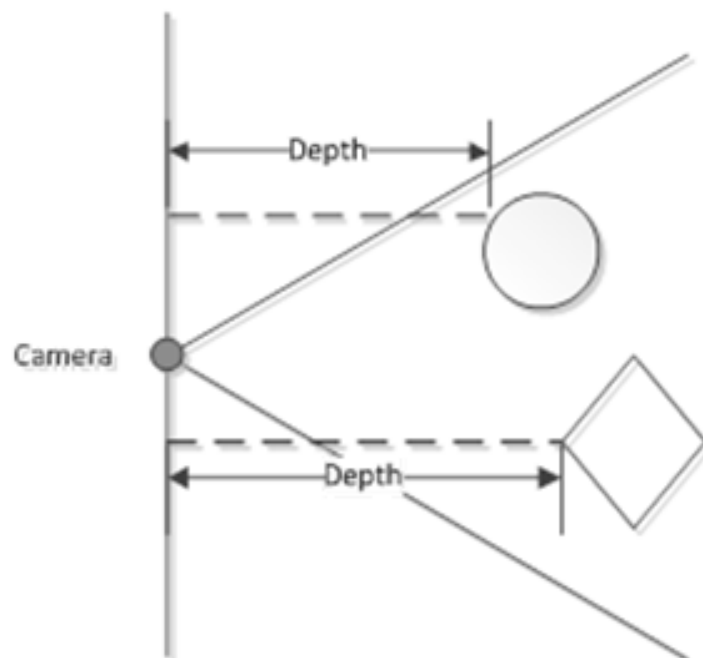
| | |
|-------------------------------|---|
| 1. Zestaw mikrofonów | Zestaw składa się z 4 mikrofonów, które wykorzystywane są przez funkcje rozpoznawania mowy. Mikrofony te posiadają funkcję filtrującą zakłócenia. |
| 2. Emiter podczerwieni | Emituje wiązkę promieni podczerwonych, które odbijając się od powierzchni zniekształcają się i następnie są odczytywane przez kamerę głębokości. |
| 3. Kamera głębokości | Analizuje zniekształconą przez obiekt wiązkę promieni podczerwonych, tworząc model 3D pomieszczenia i obiektów znajdujących się w nim. |

| | |
|---|---|
| 4. Automatyczny kontroler nachylenia | Dostosowuje położenie sensora w sposób automatyczny w zależności od rozmiarów śledzonych obiektów. |
| 5. Kabel połączeniowy | Nie jest to zwykły kabel z gniazdem USB. Z tego względu do połączenia sensora Kinect z komputerem potrzebny jest dodatkowy zasilacz z kablem USB. |
| 6. Kolorowa kamera RGB | Działa podobnie jak kamera internetowa, przesyła serie obrazów do komputera ze standardową prędkością 30 klatek na sekundę. |

Kinect tworzy obraz głębościowy poprzez projekcję specjalnego wzoru (skomponowanego z wielu małych kropek) w podczerwieni oraz odczyt zniekształceń tego wzoru przez kamerę. Są one przetwarzane na informację o odległości obiektów od sensora. Tworzenie danych o szkieletach oraz rozpoznawanie mowy również wykonywane jest po stronie urządzenia.

3. Wykonanie

Pierwszym etapem przetwarzania było pobranie chmury punktów odległości z sensora kinect. Dane te są przekazywane w postaci tablicy jednowymiarowej opisujące głębości punktów na płaszczyźnie 640x480 punktów. Każdy z tych punktów opisywany jest przez odległość w milimetrach. Sensor kinect dostarcza odległości pomiędzy płaszczyzną styczną do sensora a punktem znajdującym się przed sensorem w odległości od 0.8m do 4m (rysunek poniżej).



Następnym etapem było przetworzenie pobranej chmury punktów oraz wykrycie konkretnych obiektów na powierzchni badanej. Aby wykryć poszczególne obiekty znajdujące się przed sensorem wykorzystaliśmy różnicę w wysokościach poszczególnych pikseli. Dla danych odległości punktów x_1 oraz x_2 policzyliśmy

$$\Delta = (x_1 - x_2)$$

Gdzie x_1 oraz x_2 to odpowiednio odległość punktu x_1 oraz x_2 od sensora w milimetrach.

Generowaliśmy następnie spójne obszary, których delta nie przekraczała pewnego wyznaczonego progu. Okazało się iż najlepsze rezultaty osiągają małe delty, od 10 do 20 mm.

Generowanie tych wspólnych obszarów odbywało się poprzez przechodzenie chmury punktów algorytmem podobnym do BFS, lecz zmodyfikowanym na potrzeby Kinect'a.

Podczas pracy nad projektem odkryliśmy także, iż Kinect posiada obszary, dla których nie potrafi obliczyć odległości. Wynika to z odstępu między emiterym a kamerą głębościową – wzór z emiterym nie dociera do niektórych obszarów, które widzi kamera. Punkty te przyjmują odległość równą zero. Powodowały one spore

utrudnienia w wykrywaniu obiektów. Aby poradzić sobie z tym, pomijaliśmy te punkty, jako, że część z nich reprezentowała wadliwe piksele. Niestety liczenie średniej z sąsiednich punktów nie dało zadowalających rezultatów, obiekty były źle kolorowane.

Pseudokod algorytmu znajduje się poniżej:

Q – kolejka punktów

Dodaj do Q pierwszy punkt

Póki Q nie jest pusta

$p = Q.pop();$

dla każdego punktu pp stycznego z p

jeśli pp nie był odwiedzony

jeśli $|pp - p| < \text{delta}$

koloruj pp

dodaj pp to Q

oznacz pp jako odwiedzony

pokoloruj pierwszy punkt

4. Wnioski

- a. Urządzenie Kinect może być wykorzystane do prostego wykrywania płaszczyzn na pewnym bliskim obszarze w przestrzeni zamkniętej.
- b. Pewna niedokładność urządzenia musi być zrekompensowana (wiele punktów miało wartość zero)
- c. Dobre i wydajne przetwarzanie tablicy jednowymiarowej punktów jest kluczowa do osiągnięcia optymalnej wydajności.