

LODZ UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL, ELECTRONIC,
COMPUTER AND CONTROL ENGINEERING

PHD THESIS

**Modelling 3D scenes from sequences of depth
images in a personal micro-navigation aid
for the visually impaired**

Mateusz Owczarek

Supervisor:
prof. dr hab. inż. Paweł Strumiłło
Auxiliary supervisor:
dr inż. Piotr Skulimowski

Łódź, 2019

Abstract

The overarching goal of the thesis is to develop a method for three-dimensional scene modelling from sequences of depth images in a personal micro-navigation aid for the visually impaired. An appropriately built model will aid the visually disabled in non-visual perception of the environment, orientation and mobility in unknown environments. The underpinning idea of the model is to reliably detect the ground plane (GP) region in depth images and recursively track its orientation in a 3D scene. The detected GP region also simplifies identification of positions of other scene objects as it designates an area that is devoid of potential obstacles.

In the proposed method, the position of the GP in the camera coordinate system is estimated by a Particle filter (PF). The PF is a statistical method based on the Sequential Monte Carlo method. A fundamental advantage of this method is that it is not limited to unimodal distributions of the estimated states that are modelled by a large population of particles. The parameters associated with the particles undergo prediction and update phases in search for the statistically best estimations of the modelled process state. In the proposed application the estimated state is the orientation of the GP and its motion in consecutive depth images.

Evaluation of the proposed methods in terms of GP and obstacle detection was performed on 38 image sequences, recorded both indoor and outdoor in different lighting and weather conditions. Sequences cover a defined number of situations and all contain a large-sized chequerboard located in the scene. The position of the chequerboard expressed in the camera coordinates was recognized as a ground truth for the evaluation. Tests showed that the mean values of precision and accuracy are 97% and 98% for the GP and 81% and 99% for obstacle segmentation, respectively. The root-mean-square errors of pitch and roll rotation angles of the GP do not exceed 1.3° and 0.8° on average, respectively.

The PF algorithm was mapped onto the architecture of the graphics processor unit (GPU). After proper selection of parameters the total execution time of the PF algorithm pipeline on the tested hardware (NVidia Geforce GTX 960M with 640 CUDA cores) is approximately 43ms on average, therefore the expected frame rate exceeds 23 frames/s. Deployment of the PF onto the GPU hardware has enabled to speed up the computations by a factor of about 2.5 in comparison to computations on a general purpose multicore processor.

The proposed method for 3D scene modelling and analysis is to become a part of an electronic travel aid (ETA) built with the aim of enhancing travel and mobility of the visually impaired by non-visual, acoustic and haptic presentation of the environment.

Streszczenie

Głównym celem pracy jest opracowanie metody modelowania sceny trójwymiarowej na podstawie sekwencji obrazów głębi na potrzeby mikronawigacji dla osób niewidomych. Odpowiednio zbudowany model będzie pomocny osobom niepełnosprawnym wzrokowo w niewizualnej prezentacji otoczenia, orientacji przestrzennej i poruszaniu się w nieznanym środowisku. Podstawową własnością modelu jest wykrywanie obszaru płaszczyzny podłoża w obrazach głębi i rekursywne śledzenie jej orientacji w scenie trójwymiarowej. Wykryty obszar płaszczyzny podłoża ułatwi również identyfikację położenia innych obiektów sceny. Wynika to z faktu, że wykryta płaszczyzna podłoża jest obszarem przestrzeni pozbawionym innych obiektów sceny, np. przeszkód na drodze osoby niewidomej.

W proponowanej metodzie położenie płaszczyzny podłoża we współrzędnych kamery jest estymowane z zastosowaniem filtracji cząsteczkowej. Filtracja cząsteczkowa jest metodą statystyczną opartą na sekwencyjnej metodzie Monte Carlo. Podstawową zaletą tej metody jest brak ograniczenia związanego z warunkiem jednomodalności rozkładu prawdopodobieństwa estymowanych stanów, które są modelowane przez duży zbiór cząsteczek. Parametry opisujące cząsteczki podlegają filtrowaniu wg. iteracyjnego schematu przewidywania i aktualizacji w poszukiwaniu statystycznie najlepszych oszacowań modelowanego stanu procesu. W proponowanym rozwiązaniu estymowany stan jest związany z orientacją płaszczyzny podłoża i jej zmianami spowodowanymi ruchem płaszczyzny w kolejnych ramkach sekwencji obrazów głębi.

Ocenę proponowanych metod w zakresie wykrywania płaszczyzny podłoża oraz przeszkód przeprowadzono dla 38 sekwencji obrazów głębi, zarejestrowanych zarówno wewnątrz jak i na zewnątrz budynków oraz, w różnych warunkach oświetlenia i warunków atmosferycznych. Sekwencje obrazów zostały zarejestrowane dla określonej liczby różnych przemieszczeń układu akwizycji obrazu i zawierają dużą szachownicę testową umieszczoną w scenie. Pozytcje szachownicy wyrażone we współrzędnych kamery wykorzystano jako „złoty standard”, który posłużył do weryfikacji algorytmów wykrywania położenia płaszczyzny podłoża. Na podstawie przeprowadzonych testów pokazano, że uzyskane średnie wartości precyzji i dokładności wynoszą odpowiednio 97% i 98% dla wykrywanej płaszczyzny podłoża oraz 81% i 99% dla wykrywanych przeszkód. Wartości średnio-kwadratowe błędów dla estymacji kątów nachylenia płaszczyzny pitch oraz roll nie przekraczają odpowiednio 1.3° oraz 0.8° .

Algorytm filtracji cząsteczkowej został zaimplementowany na procesorze graficznym (GPU) z zastosowaniem technologii CUDA. Właściwy dobór parametrów algorytmu pozwolił na uzyskanie całkowitego czasu obliczeń (dla procesora GPU NVidia Geforce GTX 960M z 640 rdzeniami CUDA) o średniej wartości 43 ms, tym samym uzyskana szybkość analizy ramek obrazów głębi przekracza 23 obrazy/sek. Implementacja algorytmu filtracji cząsteczkowej na procesorze graficznym umożliwiła przyspieszenie obliczeń o około 2,5 raza w porównaniu do czasu obliczeń uzyskanych dla wielordzeniowego procesora ogólnego przeznaczenia.

Proponowana metoda modelowania i analizy scen 3D jest elementem elektronicznego systemu wspomaganie osób niewidomych w samodzielnym poruszaniu się. Wspomaganie to będzie realizowane na podstawie niewizualnej prezentacji akustyczno-haptycznej otoczenia.