

Segmentarea semantică pentru secvențe stereo

Andrei Neculai

Rezumat

Abilitatea de a percepe și de a înțelege structura lumii înconjurătoare este înăscută omului, dar pentru roboți acest lucru reprezintă o provocare. În ultimii ani abordări bazate pe reconstrucția scenei sau pe segmentare semantică au condus la rezultate din ce în ce mai bune și au arătat că aceste rezultate se pot obține și în condiții de timp real. Această lucrare prezintă o soluție de segmentare semantică pentru secvențe stereo bazată pe hărți de adâncime.

Utilizând secvențe stereo putem folosi pentru fiecare pixel, pe lângă informații de culoare și localizarea 3D a acestuia. Utilizarea corespunzătoare a acestei informații duce la rezultate mai de încredere deoarece acestea nu vor fi influențate de gradul de iluminare, de textura obiectelor, de orientarea camerei sau de schimbarea perspectivei.

Informațiile de adâncime, împreună cu informațiile de culoare din imagine, sunt folosite pentru a calcula o serie de trăsături la nivelul unui grup de pixeli numit superpixel. Calcularea trăsăturilor la nivel de superpixel oferă o serie de avantaje față de realizarea clasificării pentru fiecare pixel în parte. În primul rând, trăsăturile ce vor fi calculate în următorii pași sunt mai bine definite în raport cu o suprafață decât la nivel de pixel. În al doilea rând, cresc șansele ca segmentarea finală să aibă margini bine definite. În final, utilizarea superpixelilor aduce și un spor de performanță prin reducerea dimensionalității sistemului.

Trăsăturile alese pentru a fi calculate în această lucrare sunt: normala la suprafață, înălțimea deasupra solului, planaritatea locală, planaritățile vecinilor, distanța față de planul YOZ și culoarea. Aceste trăsături vor fi folosite pentru antrenarea unui clasificator de tip Random Forest în urma căreia va rezulta un model ce va fi folosit ulterior pentru realizarea segmentării semantice.

Pentru utilizarea componentei temporale a intrării aplicăm o tehnică de fuziune temporală ce are ca scop combinarea informațiilor cadrului curent cu informațiile acumulate din cadrele precedente. Folosind această abordare vom putea menține o stare globală a scenei ce va fi folosită pentru a reduce impactul erorilor inerente de calcul al informațiilor de adâncime.

Pentru a îmbunătăți rezultatele obținute în urma clasificării vom folosi un algoritm de optimizare de tip Graph Cut ce are ca scop eliminarea erorilor de clasificare. Acest lucru este realizat prin minimizarea unor costuri asociate fie atribuirii unei anumite clase unui superpixel, fie atribuirii de clase diferite unei perechi de superpixeli vecini. Minimizarea sumei acestor costuri va realiza un compromis între folosirea clasei rezultate în urma clasificării și păstrarea integrității obiectelor din imagine în segmentarea finală. Pentru îmbunătățirea rezultatelor vom refolosi în această etapă starea globală a scenei obținută în urma fuziunii temporale. Clasificarea realizată pentru un superpixel în cadrul curent va fi combinată cu clasificările superpixelului respectiv în cadrele precedente, combinație din care va rezulta o clasificare mai de încredere.

Aplicația a fost testată folosind două seturi de date: un set de 58 imagini provenite din suita KITTI [1], util pentru testarea utilității aplicației în domeniul mașinilor autonome și un set de 377 de imagini achiziționate dintr-un mediu virtual, din perspectiva unui pieton, pentru testarea aplicației în domeniul tehnologiilor asistive pentru persoane nevăzătoare. Rezultatele preliminare obținute sunt promițătoare și arată faptul ca această abordare poate fi folosită cu

succes în domenii precum dezvoltarea mașinilor autonome, tehnologiilor asistive pentru persoane nevăzătoare sau al navigării autonome a roboților mobili.

Cuvinte cheie: segmentare semantică, computer vision, sisteme de vedere stereo, clasificare, superpixel