

Segmentare semantică pentru secvențe stereo

Otilia Jacotă

Rezumat

Definirea problemei

Segmentarea semantică, sau clasificare pixelilor, reprezintă partiționarea imaginii în părți semnificative din punct de vedere semantic și clasificarea fiecărei părți într-o clasă predeterminată (drum, clădire, cer), prin asignarea unei etichete fiecărui pixel.

Parsarea scenei, procesul care se referă la clasificarea și segmentarea obiectelor dintr-o imagine, este una din problemele fundamentale ale domeniului computer vision. O bună parsare a scenei poate avea aplicații în diverse domenii fiind utilă pentru recunoașterea obiectelor, modelarea 3D a scenelor urbane și sistemele de asistență pentru șoferi.

Lucrarea de față se axează pe segmentarea semantică a secvențelor stereo, preluate în medii urbane și își propune studiul unor algoritmi prin care problema ar putea fi rezolvată.

Rezultate

În mod tradițional, segmentarea semantică se bazează doar pe informațiile de culoare extrase din imagini, astfel utilizarea hărților de adâncime obținute din hărțile de disparitate, aduce avantaje notabile în ceea ce privește segmentarea. În primul rând, adâncimea nu variază în funcție de lumină, textură, în al doilea rând este invariantă în raport cu poziția camerei și schimbarea perspectivei, astfel utilizarea acestor hărți de adâncime îmbunătățește procesul de segmentare. În cazul utilizării camerelor stereo, hărțile de adâncime pot fi calculate pe baza hărților de disparitate. Valorile hărții de disparitate reprezintă deplasarea pixelilor din imaginile din stânga față de imaginile din dreapta, achiziționate de camera stereo.

După obținerea hărților de adâncime, algoritmul propune, un pas de preprocesare comun tuturor algoritmilor de parsare a scenei, în care imaginea este suprasedimentată (over-segmentation) în regiuni generice, numite superpixeli. Printre avantajele suprasedimentării se numără: fiecare superpixel din spațiul 2D poate fi văzut ca o zonă compactă în spațiul 3D, trăsăturile calculate utilizând hărțile de adâncime pot fi ușor definite pentru un superpixel, suprasedimentarea poate reduce complexitatea calculelor, dacă se consideră fiecare superpixel ca un eșantion, numărul total de eșantioane este mai mic decât dacă s-ar lua în calcul fiecare pixel separat.

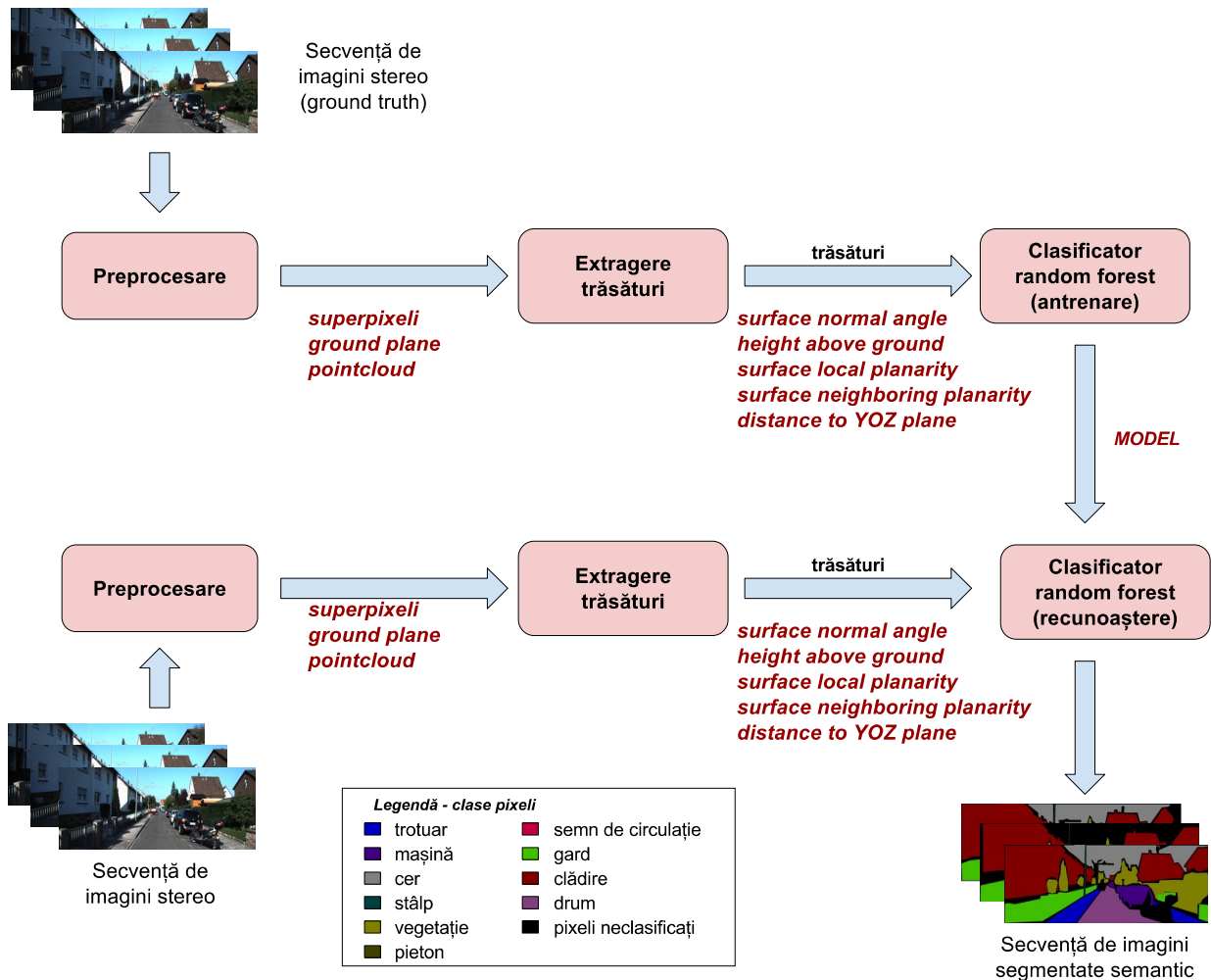
Cele cinci trăsături extrase pentru fiecare superpixel dintr-o imagine sunt: normala la suprafață (*surface normal*), înălțimea deasupra solului (*height above ground*), planaritatea locală (*surface local planarity*), planaritățile vecinilor (*surface neighboring planarity*), distanța față de planul YOZ. Aceste trăsături sunt normalizate după care sunt antrenate într-un algoritm de clasificare. Algoritmul de clasificare este bazat pe arbori aleatori (random forest), eroarea pe mulțimea de antrenare fiind foarte mică în acest caz. Algoritmul de învățare este constituit din două părți, una de antrenare ce are ca scop final obținerea unui model ce va fi ulterior aplicat pentru partea a doua, cea de recunoaștere, în urma căreia se va obține imaginea segmentată semantic.

Aplicația poate fi împărțită în următoarele module: modulul de preprocesare, modulul de calcul al trăsăturilor, modulul de antrenare și modulul de recunoaștere.

Modulul de preprocesare se ocupă de suprasedimentarea imaginii, adică împărțirea acesteia în zone uniforme numite superpixeli, determină point cloud-ul și ground plane-ul

folosite pentru ulterioarele procesări.

Modulul de extragere a trăsăturilor presupune calculul trăsăturilor, enunțate anterior, pentru fiecare superpixel determinat. Acestea sunt normalizate și antrenate într-un algoritm de învățare la finalul căruia se va obține modelul necesar modulului de recunoaștere. Pe baza acestui model și a unei secvențe de imagini stereo date la intrare, modulul de recunoaștere va clasifica fiecare pixel în clasa din care face parte. La finalul acestei procesări se va obține o secvență de imagini segmentată semantic. Arhitectura aplicației și modulele componente se pot observa în figura de mai jos.



Pentru partea de testare se vor utiliza două seturi de date. Un set de date achiziționate în mediul real, de la KITTI Vision Benchmark Suite, și un alt set de date achiziționate dintr-un mediu virtual. Seturile de date conțin perechea de imagini stereo, harta de disparitate, imaginile segmentate semantic și măștile pentru obținerea ground plane-ului, pentru fiecare imagine în parte. Testele s-au realizat cu seturi de date specifice pentru două domenii de aplicabilitate: automotive, respectiv tehnologii asistive.

Concluzii

Segmentarea semantică a imaginilor este unul din task-urile provocatoare ale domeniului de computer vision și totodată este unul dintre pașii cei mai importanți, în procesul de înțelegere a imaginilor. Soluția propusă în lucrarea de față prezintă un algoritm pentru segmentarea semantică a secvențelor stereo, bazat pe hărțile de adâncime obținute din hărțile de disparitate.

Cuvinte cheie: *segmentare semantică, hartă de adâncime, hartă de disparitate, trăsături, superpixeli, point cloud, ground plane*