

Detectarea regiunilor ce conțin text în secvențe video

Scripcariu Răzvan

Rezumat

Proiectul constă în realizarea unei aplicații pentru calculator capabilă să localizeze textul dintr-o imagine în timp real. Utilizatorul va putea alege o imagine, un film sau camera web conectată la calculator și va putea observa rezultatele algoritmului de detectare a textului în timp real. Funcționalitatea acestui tip de aplicație poate fi reliefată prin utilizarea sa în situații precum asistența persoanelor cu probleme vizuale, programarea roboților autonomi sau pentru localizarea și recunoașterea automată a clădirilor, prin interpretarea textului de pe acestea.

Principala componentă a proiectului este algoritmul de detecție a textului. Acesta cuprinde două module, cu funcționalități diferite. Primul modul aplică filtrele și algoritmii necesari pentru aflarea parametrilor de selectare a componentelor. Al doilea modul folosește rezultatele obținute pentru a elimina obiectele ce nu pot reprezenta text din imagine și grupează componentele rămase în cuvinte. Filtrele principale folosite sunt culoarea obiectelor, dimensiunile acestora și poziția în imagine. Interfața cu utilizatorul a fost concepută cât mai intuitiv posibil. Toate acțiunile se pot realiza direct din bara de meniu a aplicației.

În cazul în care imaginile preluate sunt complexe și conțin foarte multe obiecte, algoritmul de detectare a textului nu va putea prelucra fiecare cadru primit. Pentru a nu împiedica vizualizarea imaginilor preluate de la camera web sau din filmul ales de utilizator, algoritmul de detectare a textului va rula pe un fir de execuție separat. Astfel, cadrele ce nu au putut fi prelucrate vor fi afișate fără întârziere.

Pentru a obține performanțe cât mai bune, primul modul al aplicației a fost implementat folosind CUDA¹. Astfel, folosind resursele plăcii video, fiecare algoritm utilizat a fost paralelizat astfel încât pentru fiecare pixel din imagine să existe un fir de execuție ce efectuează toate calculele necesare. Totodată, acest lucru implică și necesitatea existenței unei plăci video NVIDIA² ce suportă CUDA, aceasta fiind principala limitare a aplicației. Știind că apelurile funcțiilor ce rulează pe placa video sunt asincrone din punctul de vedere al procesorului, în cazurile posibile am ordonat prelucrarea datelor astfel încât o parte din calcule să fie executate în paralel.

Obținerea unor rezultate cât mai bune în ceea ce privește detectarea textului, necesită ca imaginile pe care se va aplica algoritmul de detectare a textului să fie de o calitate cât mai bună. În cazul în care regiunile ce conțin text nu sunt clare sau anumite litere lipsesc, algoritmul nu va fi capabil să localizeze textul cu succes. Aplicația nu este dependentă de rezoluția imaginilor primite. Acestea vor fi scalate la dimensiunea interfeței încă de la preluare.

- 1 CUDA - abreviat din limba engleză „Compute Unified Device Architecture”, este o arhitectură hardware și software pentru calculul paralel al datelor.
- 2 NVIDIA - este o corporație multinațională americană specializată pe producerea de procesoare grafice.