

Detecția persoanelor/mișcării folosind camere de supraveghere

MSS01

1st Panciuc Ilie-Cosmin

Universitatea Tehnica "Gheorghe Asachi" Iasi
Iasi, Romania
ilie-cosmin.panciuc@student.tuiasi.ro

2nd Anghel Ioana

Universitatea Tehnica "Gheorghe Asachi" Iasi
Iasi, Romania
ioana.anghel@student.tuiasi.ro

Abstract—Identificarea persoanelor este una dintre cele mai mari provocări din Computer Vision și a fost studiat de la începutul prelucrării de imagine digitale.

Index Terms—Computer Vision, human detection, opencv

I. INTRODUCTION

În era contemporană, securitatea este o prioritate esențială, iar camerele de supraveghere video joacă un rol esențial în menținerea acesteia. Într-un peisaj digital în continuă evoluție, unde volume semnificative de date sunt generate de camerele de securitate, eficacitatea în identificarea și urmărirea indivizilor devine crucială.

Proiectul curent își propune să construiască o aplicație robustă și eficientă care să poată procesa datele furnizate de aceste dispozitive video cu scopul identificării umane în timp real.

II. STATE-OF-THE-ART

În dezvoltarea acestei aplicații am consultat o varietate de lucrări SOTA dintre care cele mai notabile sunt YOLO(You Only Look Once), Faster R-CNN (Region-based Convolutional Neural Network) și SSD (Single Shot Multibox Detector)

A. YOLO

Această soluție a fost dezvoltată pentru a realiza detecția în timp real a peste 9000 de categorii de obiecte. YOLO aduce o abordare inovatoare în detecția obiectelor prin tehnică să de parcurgere a rețelei o singură dată . Eficientă în timp real și precizia fac YOLO un candidat ideal pentru evaluarea performanței aplicației noastre.

B. Faster R-CNN

Arhitectură Faster R-CNN este o combinație între o rețea neuronală convolutională și o rețea de regiuni. Abordarea inovativă a Faster R-CNN constă în implementarea unui Region Proposal Network(RPN) acesta rețea identifică în mod eficient potențiale obiecte din imaginea care mai apoi să propună o împărțire a imaginii în secțiuni.

C. SSD

La parcurgerea imaginii SSD creează propuneri de obiecte pe care mai apoi le evaluează bazându-se pe "ancorele" plasate anterior abordare care măresc semnificativ viteza de procesare și acuratețea . Spre deosebire de aplicațiile prezentate anterior SSD folosește o tehnică de "ancorare". SSD este o altă arhitectură care abordează metoda de parcurgere a rețelei într-un singur pas .

D. Cascade R-CNN

Această abordare este o extensie în mai mulți pași a R-CNN care se focusează pe identificarea mai exactă și dar mai încheată decât Fast R-CNN. Soluția această conduce la o rată mai mică a fals pozitivelor.

III. RELATED WORKS

În dezvoltarea acestei aplicații am studiat o multitudine de soluții prezente pe piață dintr-o varietate de domenii și implementări .

- [1] Această lucrare prezintă o abordare cadru semantic pentru detectarea umană care utilizează structuri ierarhice și partajarea caracteristicilor. Abordarea se bazează pe trei niveluri de caracteristici semantice: cuvânt, propoziție și paragrafe. Poate gestiona ocluziile și detectarea obiectelor non-rigide. Experimentele demonstrează eficacitatea și eficiența în detectarea pietonilor și în detectarea umană articulată. Autorii discută, de asemenea, îmbunătățirile viitoare, cum ar fi învățarea automată a cuvintelor și propozițiilor și extinderea la alte probleme de detectare a obiectelor.
- [2] Lucrarea prezintă o abordare similară cu cea pe care am ales să o implementăm, autorii propun detecția persoanelor utilizând un filtru de contur, astfel reușind să obținem o procesare de imagine la 20 de frame-uri pe secundă care este la același nivel de acuratețe cu alte opțiuni SOTA.
- [3] Acest articol prezintă o abordare extrem de simplă a detecției umane folosind un filtru de corelare al magnitudinii. Cheia abordării prezentate constă în abordarea antrenării modelului, această fiind realizată prin folosirea

unui ASEF(average of synthetic exact filters) fapt care permite procesarea imaginilor la peste 25 de frame-uri pe secundă cu o acuratețe de peste 94.7% cu mai puțin de o alarmă falsă per imagine pentru imagini cu mulțimi.

IV. METHOD DESCRIPTION

În urmă consultării mai multor articole am luat decizia de a implementa o abordare în doi pași care în ciuda constrângerilor hardware oferă o acuratețe satisfăcătoare în contextul unui flux constant de imagini cu o rezoluție mică.

Componentă software pe care o vom utiliza în acesta soluție este algoritmul HOG(Histogram of oriented gradients) implementat în biblioteca open source opencv. În ciuda limitărilor de care da dovadă algoritmul ne oferă o flexibilitate și un potențial de optimizare care îl transformă într-un favorit în fața celorlalte metode.

A. Prelucrarea frame-ului

Pentru a optimiza și a masca o parte din problemele algoritmului vom face o serie de prelucrări asupra imaginii.

- Prima dintre aceste prelucrări este scalarea imaginii la o rezoluție cu o lățime a imaginii de 640px. Am ales această dimensiune în urmă unei multitudini de teset și în raport cu o serie de parametri ai algoritmului HOG pe care îi vom discuta în secțiunea următoare. Operația de scalare a imaginii va fi implementată prin componentă resize a modului imutils. Am ales să utilizăm acest modul în ciuda existenței unui modul similar din biblioteca opencv datorită capacității sale de a păstra aspectul imaginii originale, fapt care îmbunătățește performanță algoritmului.
- Cea de a doua prelucrare pe care o vom aplica frame-ului este grayscale-ul, această operație îmbunătățește semnificativ performanță algoritmului limitând spectrul de culoare pe care un bit îl poate avea.

B. Identificarea persoanelor

A doua operație a programului constă în identificarea efectivă a persoanelor. Funcția de detecție pe care implementarea algoritmului HOG ne-o pune la dispoziție ne oferă o libertate ideală pentru optimizarea algoritmului relativ la nevoile care apar în prelucrarea imaginilor cu diverse nivele de calitate și iluminare. Cele mai relevante dintre aceste opțiuni sunt reprezentate de argumentele winStride care determina de câte "ferestre" este nevoie pentru a identifica un obiect, în cazul aplicației noastre valoarea tuplei (8,8) oferă rezultate satisfăcătoare. Un alt parametru relevant pentru optimizarea algoritmului îl reprezintă opțiunea scale care determina în câte nivele este scalată imaginea noastră, acest parametru are un efect semnificativ asupra performanței întrucât mărește semnificativ timpul de calcul al algoritmului.

V. PRELIMINARY RESULTS

În urmă testării aplicației pe o varietate de dataset-uri precum "Penn-Fudan", VIRAT și un dataset propriu am obținut o acuratețe variată în funcție de calitatea și luminozitatea imaginii.

- În condiții de iluminare bună și rezoluție mare am obținut o acuratețe de peste 93.4%

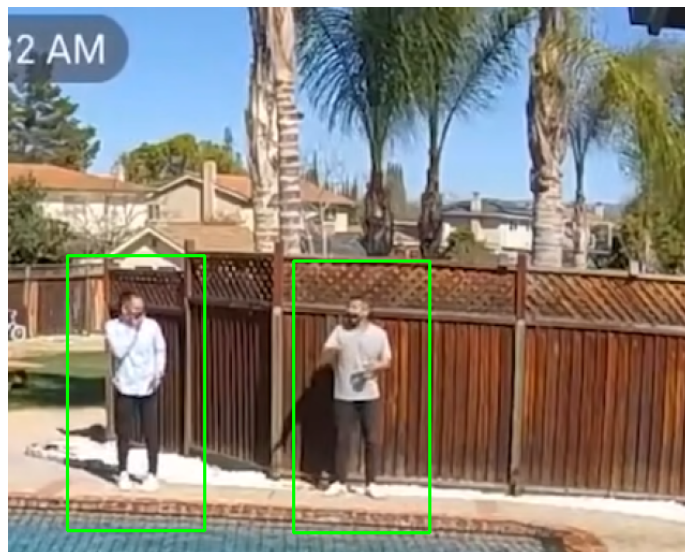


Fig 1. Identificare în condiții ideale de lumina

- Când luminozitatea scade dar calitatea rămâne mare reușim să obținem o acuratețe de 89.1% acest

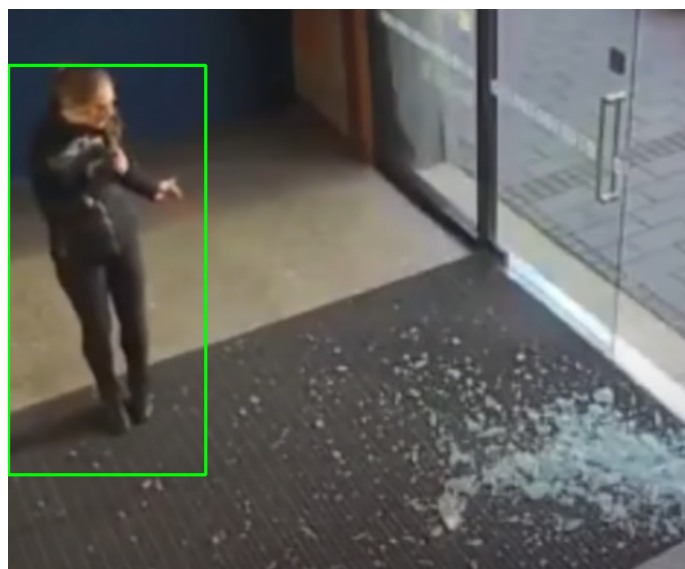


Fig 2. Identificare în condiții reduse de lumina

- Atunci când calitatea imaginii scade acuratețea pe care reușim să o obținem este aproape de 85.2%
- Cele mai proaste rezultate le-am obținut în condiții de iluminare proastă și calitate a imaginii scăzute, acuratețea pe care am reușit să o obținem în această situație este de doar 80.1%

VI. PRELIMINARY CONCLUSIONS

În ciuda limitărilor curente legate de frame rateul și calitatea videoclipurilor suntem încrezători că aplicația poate fi

optimizată mai departe. Deși am ales să nu implementăm o rețea neuronală am reușit să obținem rezultate satisfacatoare raportat la soluțiile prezente la momentul actual pe piață.

REFERENCES

- [1] Wahyu Rahmانيar, Ari Hernawan, 'Real-Time Human Detection Using Deep Learning on Embedded Platforms: A Review', National Central University Taiwan, "<https://journal.umy.ac.id/index.php/jrc/article/view/10558>"
- [2] Jianxin Wu, Christopher Geyer, James M. Rehg, Real-time human detection using contour cues, Shanghai, China. 09-13 May 2011, "<https://ieeexplore.ieee.org/abstract/document/5980437>"
- [3] David S. Bolme, Yui Man Lui, Bruce A. Draper, J. Ross Beveridge, "Simple real-time human detection using a single correlation filter", Snowbird, UT, USA , 07-09 December 2009