

# Detecția persoanelor/mișcării folosind camere de supraveghere

MSS01

1<sup>st</sup> Panciuc Ilie-Cosmin

Universitatea Tehnica "Gheorghe Asachi" Iasi

Iasi, Romania

ilie-cosmin.panciuc@student.tuiasi.ro

2<sup>nd</sup> 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 digitală.

**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 . Eficiența în timp real și precizia fac YOLO un candidat ideal pentru evaluarea performanței aplicației noastre.

### B. Faster R-CNN

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

### C. SSD

La parcurgerea imaginii SSD crează propuneri de obiecte pe care mai apoi le evaluează bazându-se pe "ancorele" plasate anterior, abordare care mărește 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, dar mai încheată decât Fast R-CNN. Soluția aceasta conduce la o rată mai mică a rezultatelor fals pozitive.

## 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, dar 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.
- [4] Lucrarea aceasta prezintă o abordare a problemei analizei CCTV folosind o rețea de tip CNN, aceasta lucrare ne-a inspirat și ne-a îndrumat spre alegerea unor dataset-uri ideale pentru testarea și verificarea aplicației noastre.

#### 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.

Componenta 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 dă dovadă, algoritmul ne oferă o flexibilitate și un potențial de optimizare pe care îl transformă într-un favorit în fața celorlalte metode.

##### A. Prelucrarea frame-ului

Pentru a optimiza și 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 urma unei multitudini de teste 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 componenta 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 performanța algoritmului.
- A doua prelucrare pe care o vom aplica frame-ului este grayscale-ul. Această operație îmbunătățește semnificativ performanța algoritmului limitând spectrul de culoare pe care un bit îl poate avea.



Fig 1. Black and white

##### 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 determină 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 determină în câte nivele este scalată imaginea noastră, acest parametru are un efect semnificativ asupra performanței întrucât mărește semnificativ timpul de calcul al algoritmului.



Fig 2. HOG processing

##### C. Urmărirea persoanelor

Cea de a treia operație pe care programul o realizează este cea de object tracking, librăria opencv ne pune la dispoziție

o varietate de trackere deja implementate. Criteriile pe care le-am utilizat în evaluarea acestora sunt viteză de procesare și acuratețea cu care acesta reușește să urmărească obiectul în diferite condiții de calitate și iluminare. În căutarea celui mai bun rezultat am testat și analizat 8 dintre cei mai populari algoritmi de urmărire puși la dispoziție de opencv.

- **BOOSTING Tracker** (boosted cascade of simple features):  
Acest algoritm este recunoscut pentru viteza sa în condiții ideale. Din păcate, chiar și în aceste condiții acuratețea algoritmului lasă de dorit, când condițiile de lumina și calitate scad, acuratețea scade și ea sub un prag minim acceptabil.
- **MIL Tracker:**  
Comparativ cu BOOSTING Tracker acest algoritm oferă o creștere sesizabilă în defavoarea vitezei de procesare, acest fapt descalifică algoritmul, fiind direct în contradicție cu scopul nostru de a aduce programul la o procesare în timp real.
- **MedianFlow Tracker;** Acest algoritm a obținut rezultate medii atât în viteză de procesare cât și în acuratețea cu care urmărește obiectele.
- **MOSSE Tracker:** De cele mai multe ori cel mai rapid algoritm de urmărire din opencv MOSSE este o varianta pe care am avut-o în vizor în momentul proiectării programului, însă acuratețea scăzută și numărul mare de erori au descalficat algoritmul în etapa finală de proiectare.
- **TLD Tracker:** La spectrul opus algoritmului MOSSE se află TLD. Acest algoritm oferă o acuratețe impresionantă, însă durata foarte mare de procesare a condus la ignorarea acetui algoritm în contextul aplicației noastre.
- **CSRT Tracker:** Unul dintre finaliștii etapei de proiectare, CSRT este o alternativă puțin mai înceată decât algoritmul pe care am ales să îl folosim. Din păcate, în contextul aplicației noastre am concluzionat că acuratețea câștigată nu este suficient de semnificativă pentru a justifica durata mai mare de procesare.
- **KCF Tracker:** Algoritmul pe care am ales să îl utilizăm în programul nostru KCF oferă un balans ideal între durata de procesare și acuratețea de urmărire, devenind algoritmul ideal pentru aplicația noastră care dorește să proceseze imaginile în timp real. În ciuda acestor beneficii algoritmul nu este perfect, având probleme de acuratețe în special atunci când obiectul urmărit este obstrucționat total.

## V. RESULTS

În urma testării aplicației de la etapa intermediară în diferite condiții de calitate și iluminare am obținut rezultate

mulțumitoare din punct de vedere al acurateții în schimb timpul de procesare al acestor imagini în contextul utilizării librăriei opencv fără posibilitatea accelerării prin intermediul GPU-ului a fost dezamăgitor. Astfel am ales să utilizăm un algoritm de urmărire al obiectelor care ar duce programul mai aproape de țelul sau de procesare real time fără să sacrifice prea mult din acuratețe.

Tabelele de mai jos prezintă rezultatele comparative ale aplicației noastre înainte de utilizarea object trackerului.

Tabel de acuratete		
Conditile imaginii	Acuratete HOG simplu	Acuratete HOG+KCF
Ideale	88.4%	83.6%
Luminozitate scazuta	84.3%	81.2%
Calitate scazuta	82.5%	80.1%
Calitate si luminozitate scazuta	80.9%	79.6%

Tabela de mai jos descrie timpul mediu de procesare al unui videoclip de aproximativ 30 de secunde care poate condține între 0 și 20 de persoane.

Timp de procesare		
Conditile imaginii	Timp de procesare HOG	Timp de procesare HOG+KCF
Ideale	2.5 minute	1.7 minute
Luminozitate scazuta	3.2 minute	2.2 minute
Calitate scazuta	2.1 minute	1.3 minute
Calitate si luminozitate scazuta	2.2 minute	1.5 minute

**\*conditii ideale 1920x1080 25 fps**

**\*conditii scazute 720p 25 fps**

De asemenea putem remarca o creștere în timpul de procesare și o scădere a acurateții cu cât numărul de persoane din imagine crește, acest fapt fiind datorat necesității de creare a unui Tracker separat pentru fiecare obiect identificat.

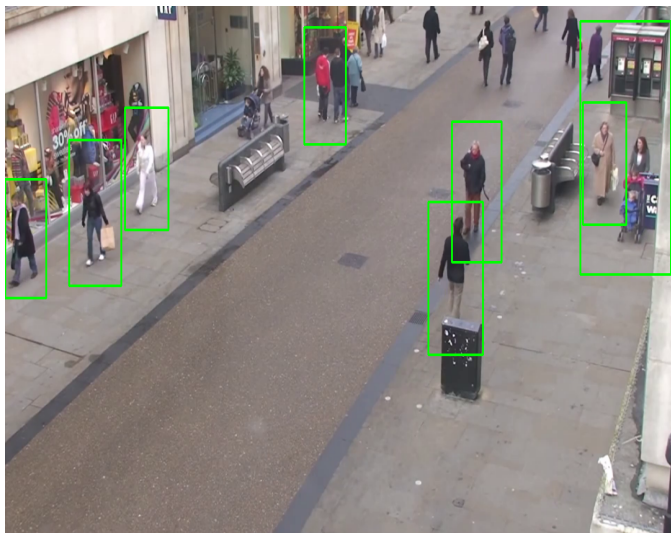


Fig 3. Identificare in conditii ideale de lumina si calitate

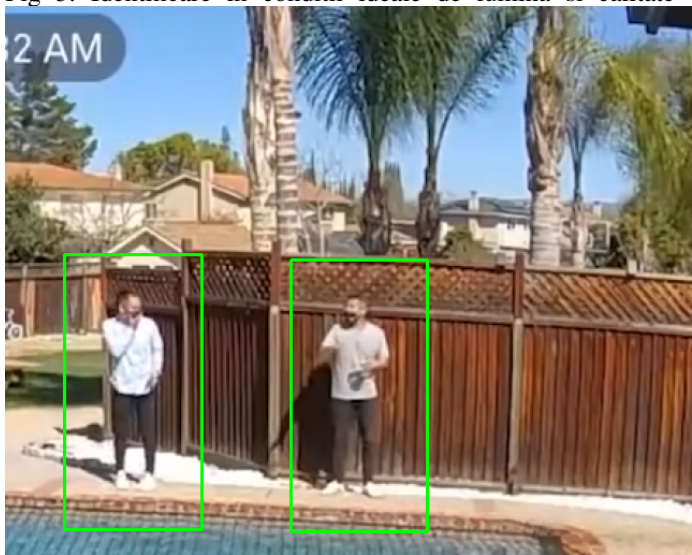


Fig 4. Identificare in conditii ideale de lumina

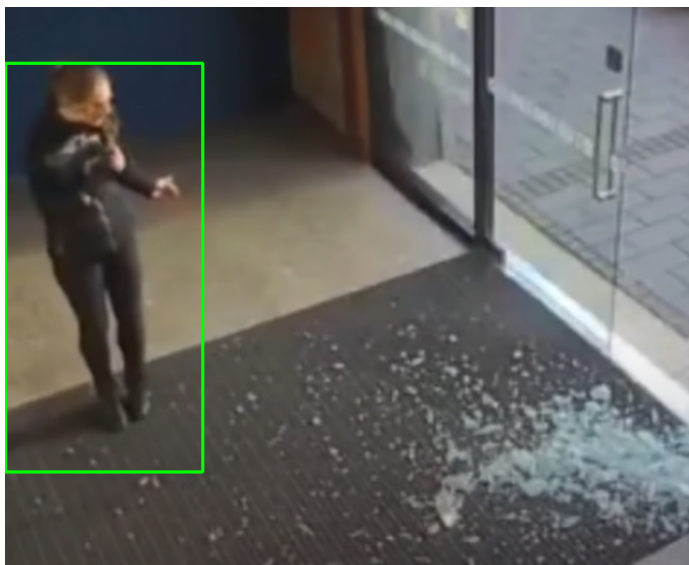


Fig 5. Identificare in conditii reduse de lumina Pentru testarea aplicatiiei am utilizat o varietate de dataset-uri precum VIRAT si Oxford Robotics de pe KAGGLE.

## VI. CONCLUSIONS

Pe parcursul acestui proiect am dobândit o varietate de cunoștințe atât din domeniul computer vision cât și din domenii conexe. Aplicația pe care am reușit să o realizăm din păcate nu a reușit să ajungă la stadiul de procesare real time spre care tinteam, în ciuda acestui fapt considerăm că am obținut un rezultat mulțumitor.

Considerăm că rezultatele aplicației pot fi îmbunătățite prin upgradarea hardware-ului pe care programul este rulat și prin utilizarea unor librării care folosesc mai mult GPU.

Suntem recunoscători și pentru aptitudinile de comunicare și lucru în echipă pe care le-am dezvoltat pe parcursul realizării acestui proiect și sperăm să dezvoltăm o soluție mai bună în viitor.

## REFERENCES

- [1] Wahyu Rahmiani, 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
- [4] Debaditya Acharya, Kourosh Khoshelham, Stephan Winter, Real-time detection and tracking of pedestrians in CCTV images using a deep convolutional neural network, Infrastructure Engineering, The University of Melbourne