

Pedestrian Detection Using Vision-Based Techniques

Ionela RUSU, Ștefan-Gheorghe PENTIUC

Abstract—In this paper we intend to present a pedestrian detection using vision based techniques.

Human posture detection process starts building a training set of images that will contain a huge number of images of pedestrian postures, called positive image and a number of images that contain only backgrounds, called negative images.

The next step is extracting the Haar-like features from image database and training them in a cascade of "weak" classifiers with AdaBoost algorithm in order to build a "strong" classifier.

Finally, we will search possible occurrence of pedestrians applying template matching techniques.

Also, are presented a number of issues that arise in pedestrian detection and tracking, constraints on the variety of backgrounds, occlusions due to accessories or self occlusions.

Index Terms—AdaBoost, computer vision, haar-like, pedestrian

I. INTRODUCERE

A cât detecția cât și urmărirea trecătorilor reprezintă provocări în viziunea artificială, mai ales în scene complexe din lumea reală, scene care de obicei implică mai multe persoane, ocluzii complicate, precum și fundaluri complexe.

Segmentarea robustă și urmărirea trecătorilor în condiții fără restricții introduce o multitudine de factori care complică și care reprezintă una din problemele cele mai dificile în Computer Vision. Acești factori care complică trebuie să fie recunoscuți și abordați de către sistemele de viziune în cazul în care detectarea trecătorilor este posibilă în scenarii reale.

Un aport deosebit în domeniul recunoașterii formelor l-au adus Paul Viola și Michael Jones, în 2001, care au realizat prima implementare de detecție în timp real a feței folosind seturi de caracteristici de tip Haar. Aceasta tehnică este disponibilă în cadrul proiectului Open CV (Intel) în format open source.

Având în vedere succesul obținut în dezvoltarea algoritmilor de detectare a feței, detectarea posturii umane devine următorul subiect de interes pentru cercetare datorită aplicabilității sale la scară largă.

Astfel, comparând tehnicile de detectare a feței cu cele de detectare a posturii umane, acestea din urmă au un grad mai mare de dificultate datorită problemelor ce apar în analiza scenelor complexe din lumea reală, scene care de obicei implică mai multe persoane, ocluzii complicate sau fundaluri încărcate.

Astfel, în cadrul tehnicilor de detectare a trecătorilor principalele probleme sunt:

Aspectul trecătorului. O mare variabilitate în aparența locală și globală a trecătorilor poate fi cauzată de diferite tipuri și stiluri de îmbrăcăminte. Prin urmare, o tehnică ce se bazează pe o anumită caracteristică, cum ar fi aspectul picioarelor, este posibil să nu poată fi aplicată pentru toate clasele de pietoni, cum ar fi cele purta o fustă.

Expunerea trecătorului. Un trecător nu este un corp rigid, din acest motiv forma va trece printr-o gamă largă de transformări datorită varietății de poziții posibile care pot fi asumate.

Orientarea trecătorului. Un trecător poate fi privit dintr-o varietate de orientări posibile în ceea ce privește imaginea camerei foto-video. De exemplu, pietonul se poate confrunta direct cu camera foto-video, adică la 0 grade sau să fie paralel cu aceasta, adică la ± 90 de grade.

Poziția trecătorului. Un trecător poate fi amplasat într-o scenă, la diferite distanțe de aparatul de fotografiat. Apariția trecătorului aproape de camera foto-video pot diferi semnificativ de cele la o distanță mai mare.

Ocluzia proprie. Silueta unui trecător poate fi, de asemenea, perturbată de o multitudine de accesorii, cum ar fi rucsaci, pălării, genti sau pungi de cumpărături.

Ocluzia în grup. Un trecător poate fi acoperit de către unul sau mai mulți pietoni sau obiecte, mai ales în cazul în care trecătorul este situat într-o mulțime.

Viziunea camerei foto. Un trecător poate fi doar parțial în domeniul de vedere al aparatului de fotografiat. Acest lucru poate, sau poate nu, indica faptul că este un trecător care intră sau iese din scenă. Acest lucru poate apărea în mai multe zone în cadrul scenei.

Obiecte non-pedestru. Nu toate obiectele prim-plan într-o scenă dată pot fi trecători, chiar dacă forma și dimensiunea sunt similare. Este dificil de prevăzut și de elaborat modele pentru toate celelalte obiecte din lumea reală care pot să apară într-o scenă.

Condițiile de mediu. Detectarea trecătorilor în scenarii în timp real aduce dificultăți suplimentare în mișcare, inclusiv fundaluri, schimbarea de luminizitate și de condițiile meteorologice, reflecții pe ferestre și umbre exprimate de trecători și a altor obiecte de prim-plan.

Variația intensității. Poate exista o lipsă de variabilitate în intensitatea de culoare între trecători, fundal și alte obiecte de prim-plan, aceasta ducând la o segmentare dificilă a imaginii.

Aspectele prezentate până în acest moment reprezintă provocări semnificative pentru tehnicile de detectare a trecătorilor în cadre unice. Cu toate acestea, obiectivul celor

mai multe aplicații de viziune artificială, cum ar fi cele utilizate pentru securitate, solicita informații nu numai despre unde sunt oamenii la un moment dat în timp, dar și ceea ce fac acești oameni. Informațiile obținute la un moment dat de timp nu oferă astfel de informații. Mai multe informații utile despre traiectoria trecătorilor pot fi obținute în cazul în care persoanele sunt urmărite pe parcursul mai multor scene.

Cu toate acestea urmărirea trecătorilor introduce noi provocări:

Mișcarea pietonilor. Un pieton poate merge într-un mod non-linear și imprevizibil, ceea ce duce la o urmărire dificilă a acestuia. De exemplu, în scenariile lumii reale un pieton se poate plimba, alerga, opri sau întoarce în mod neașteptat.

Ocluzia. Unul sau mai multe traiectorii pot fuziona într-una singură. În funcție de tehnica de urmărire, atunci când două sau mai multe traiectorii se suprapun, este posibil ca una dintre ele să fie pierdută temporar. Algoritmul de urmărire trebuie să fie capabil să ia în considerare acest lucru, astfel încât, după suprapunere, fiecare traiectorie să mențină obiectul adecvat (cel urmărit înainte de ocluzie).

Divizarea. O traiectorie poate împărțită în două sau mai multe părți. Acest lucru poate fi din cauza segmentării sărace în cadrulul actual, segmentarea săracă în frame-urile anterioare, sau, eventual, ca urmare a introducerii unui obiect în scenă.

Noi traiectorii. Trecătorii care intră în scenă trebuie să fie recunoscuți și nu confundați cu pietonii urmăriți anterior.

Traietorii pierdute. O traiectorie se pot pierde complet, din cauza zgomotului, segmentării sărace sau ieșirii obiectului de interes din domeniul de vizibilitate al camerei foto-video. Această traiectorie pierdută ar trebui să fie marcată ca inactivă și nu confundată cu alte traiectorii.

II. CARACTERISTICILE HAAR ȘI ALGORITMUL DE CLASIFICARE ADABOOST

Un proces de recunoaștere poate fi mult mai eficient în cazul în care aceasta se bazează pe detectarea de caracteristici care codifica unele informații despre clasa ce urmează a fi detectată. Acesta este cazul caracteristicilor Haar care codifica existența unor contraste orientate între regiunile din imagine.

Caracteristicile de tip Haar sunt caracteristici simple, dreptunghiulare care calculează diferența dintre sumele pixelilor din regiunile dreptunghiulare albe și negre la diferite scări, orientări și rapoarte lățime/înălțime. În figura 1 sunt prezentate cele trei tipuri de caracteristici Haar utilizate în învățarea caracteristicilor umane.

Aceste caracteristici se calculează astfel:

- caracteristica cu 2 regiuni dreptunghiulare (orizontal / vertical): calculează diferența dintre sumele de pixeli din cele două regiuni dreptunghiulare;
- caracteristica cu 3 regiuni dreptunghiulare: calculează suma pixelilor celor 2 dreptunghiuri din extremități, apoi scade suma pixelilor din dreptunghiul central.

Un set de aceste caracteristici pot fi folosite pentru a codifica contrastelor expuse de către o postura umană și

relațiile spațiale ale acestora. Caracteristicile Haar sunt denumite în acest mod deoarece acestea sunt calculate similar cu coeficienții din transformările bazate pe unda Haar.

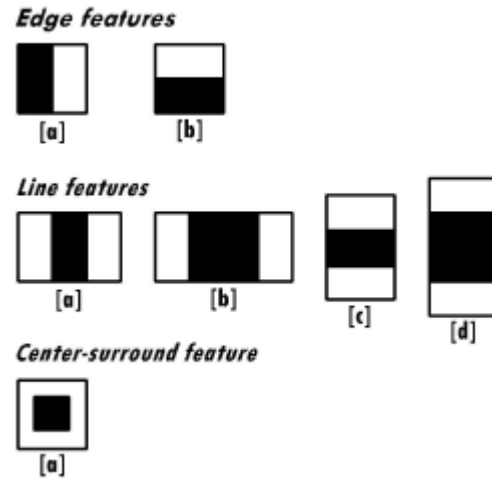


Fig.1. Prototipul caracteristicilor de tip Haar folosite în detecția posturii umane.[10]

Un clasificator ce se bazează pe caracteristici Haar, este antrenat cu câteva sute de modele ale obiectului de detectat (în acest caz, diferite vederi ale posturii umane), numite exemple pozitive, care sunt scalate la aceeași dimensiune (ex. 24x24), precum și modele negative care sunt reprezentate de imagini arbitrare, de asemenea de aceeași dimensiune.

După ce clasificatorul este antrenat, acesta poate fi aplicat într-o regiune de interes (de aceeași dimensiune ca în timpul formării) a unei imagini de intrare. Clasificatorul întoarce "1" la ieșire în cazul în care regiunea pare asemănătoare cu obiectul și "0" în sens contrar. Pentru a căuta un obiect în întreaga imagine se va muta fereastra de căutare pe toată suprafața imaginii și se va verifica fiecare locație folosind clasificatorul.

Clasificatorul este conceput astfel încât să poată fi ușor "redimensionat" pentru a putea găsi obiectele de interes la diferite dimensiuni, metoda mai eficientă decât redimensionarea imaginii în sine. Deci, pentru a găsi un obiect de dimensiuni necunoscute din imagine, procedura de scanare ar trebui să fie făcută de mai multe ori la scară diferită. Cuvântul "cascadă" în numele clasificatorului înseamnă că clasificatorul rezultat este format din mai mulți clasificatori mai simpli (etape), care sunt aplicați ulterior la o regiune de interes, până când candidatul este respins la una din etape sau a trecut de toate etapele, secvența de clasificatori $C_1, C_2, \dots, C_k$ se construindu-se astfel:

C_1 - este un clasificator construit în mod obișnuit pe setul de antrenare T

Pentru $i=2:k$

- Pentru fiecare instanță din T se atribuie o pondere proporțională cu numărul de clasificatori $C_1, \dots, C_{i-1}$ care clasifică greșit instanța respectivă
- Se re-eșantionează datele folosindu-se ponderile
- Se construiește următorul clasificator pe setul de antrenament reeșantionat

Cuvântul "boost (amplificare)" înseamnă că clasificatorii din fiecare etapă a cascadei sunt mai complecși și sunt construiți din clasificatori de bază utilizând una din cele patru diferite tehnici de amplificare/creștere (vot ponderat): discrete Adaboost, Real Adaboost, light Adaboost și Logitboost.

Clasificatori de bază sunt arbori de decizie cu cel puțin 2 frunze. Caracteristicile Haar reprezintă intrările pentru clasificatorii de bază. Caracteristicile utilizate într-un clasificator specifică forma, poziția în regiunea de interes și scara.

III. SISTEMUL DE DETECȚIE AL POSTURII UMANE

Unul dintre cele mai importante lucruri înainte de a construi clasificatorul ce va detecta trecătorii este selecția caracteristicilor care sunt reprezentative și capabile să fie aplicate în scene din lumea reală. Pentru implementarea sistemului de detecție a pietonilor am utilizat aceeași arhitectură ca și arhitectura propusă de Viola-Jones [1] pentru detecția feței.

Pentru o imagine de antrenare, avem o mulțime de caracteristici de tip Haar, noi, însă, trebuind să alegem cel mai bun set de caracteristici care pot reprezenta pietoni.

Astfel, pentru un anumit set de caracteristici și un set de imagini pozitive (pietoni), respectiv negative (non-pieton), algoritmul AdaBoost va alege cel mai bun clasificator slab în fiecare rundă. Cel mai bun clasificator slab este clasificatorul care atinge cea mai mică eroare de formare. După alegerea celui mai bun clasificator slab, AdaBoost va ajusta ponderea fiecărei imagini din setul de antrenare.

În figura 2 este prezentată arhitectura sistemului de detecție a trecătorilor.

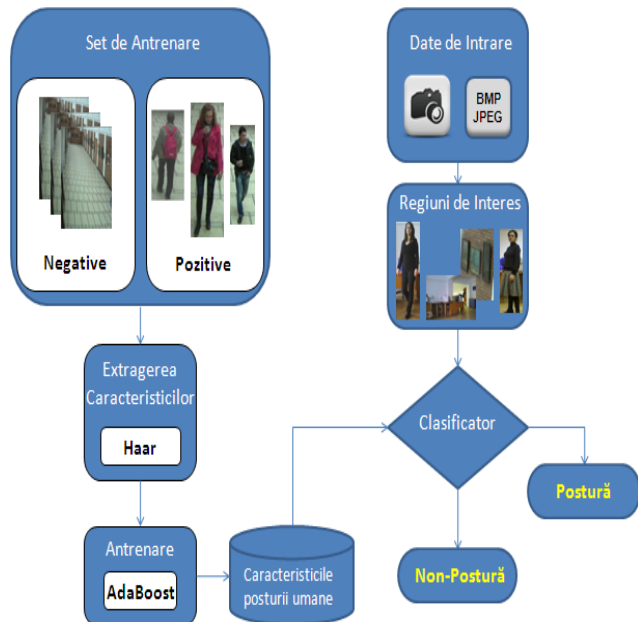


Fig. 2. Schema sistemului de detecție a posturii umane.

Procesul de detecție a posturii umane pornește de la construirea unui set de imagini de antrenare. Setul de învățare este o bază de date cu un număr foarte mare de exemple, de

aceiași dimensiune, de posturi umane, numite imagini pozitive, cât și un număr de imagini ce nu conțin posturi umane, numite imagini negative. În figura 3 este prezentat un subset de imagini pozitive (figura 3a) și un subset de imagini negative (figura 3b) din setul de antrenare folosit.

Evaluarea a zeci de mii de caracteristici asociate pentru fiecare imagine este ineficientă ca timp de calcul, de aceea se presupune că doar un număr din aceste ferestre de caracteristici pot fi clasificatori. Problema constă în găsirea acestor caracteristici și antrenarea clasificatorului. Pentru aceasta am folosit algoritmul AdaBoost. Algoritmul AdaBoost va stabili caracteristica de tip Haar folosită.



Fig. 3. Imagini folosite în setul de antrenare: imagini pozitive(a), negative(b).

Dacă în timpul analizei unei zone din imagine, nu este îndeplinită condiția unuia dintre filtrele clasificatorului Haar, acea regiune este marcată ca neconținând o postură. Fiecare regiune a imaginii va fi trecută prin întreg lanțul de filtre pentru a se stabili dacă reprezintă o postură sau nu. Regiunile care trec de toate filtrele sunt marcate ca fiind posturi.

O dată ce avem construit clasificatorul putem trece la detecția posturilor umane. După ce clasificatorul este antrenat, acesta poate fi aplicat într-o regiune de interes (de aceeași dimensiune ca în timpul formării) a unei imagini de intrare. Clasificatorul întoarce postură sau non-postură la ieșire. Pentru a căuta un obiect în întreaga imagine se va muta fereastra de căutare pe toată suprafața imaginii și se va verifica fiecare locație folosind clasificatorul.

Clasificatorul este conceput astfel încât să poată fi ușor "redimensionat" pentru a putea găsi obiectele de interes la diferite dimensiuni, metoda mai eficientă decât redimensionarea imaginii în sine. Deci, pentru a găsi un obiect de dimensiuni necunoscute din imagine, procedura de scanare ar trebui să fie făcută de mai multe ori la scară diferită.

IV. REZULTATE

Figurile 4-9 reprezintă rezultatele obținute de sistemul de recunoaștere a trecătorilor în scene cu una sau mai multe persoane.



Fig. 4. Scenă cu o singură persoană – față.



Fig. 5. Scenă cu o singură persoană – lateral.



Fig. 6. Scenă cu o singură persoană – spate.

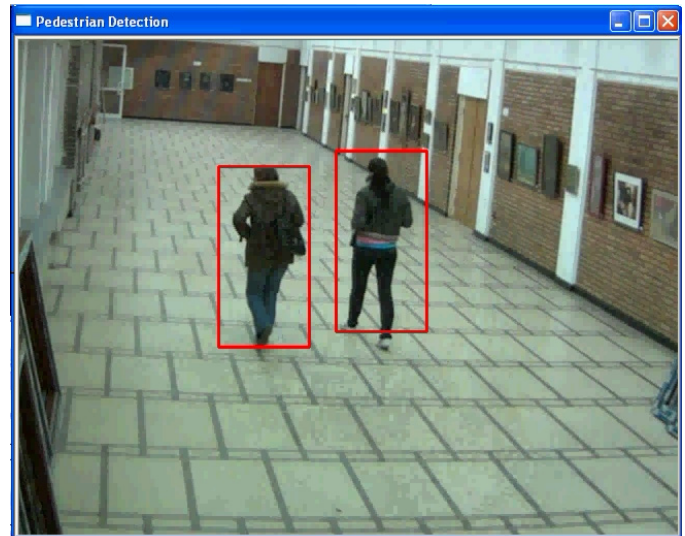


Fig. 7. Scenă cu două persoane.

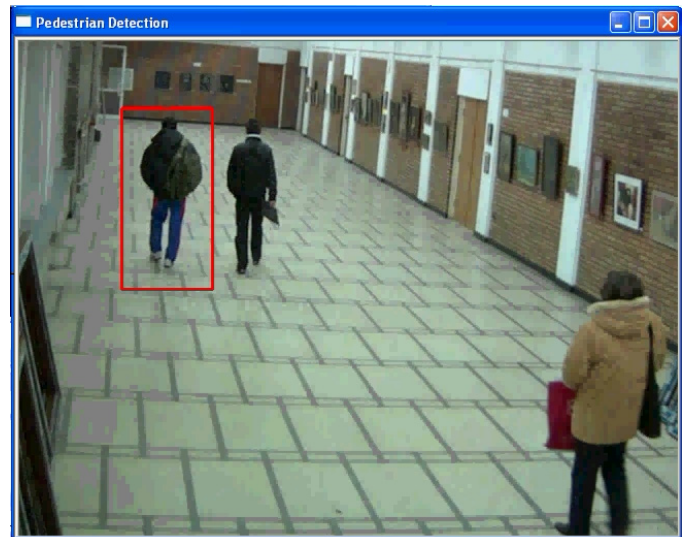


Fig. 8. Scenă cu mai multe persoane – detecție parțială.

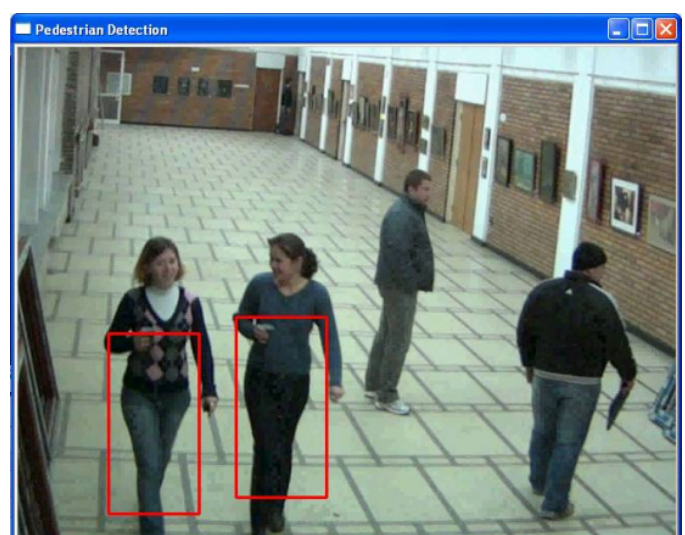


Fig. 9. Scenă cu mai multe persoane – detecție parțială.

V. CONCLUZII

Sistemul de detectare a pietonilor, fapt demonstrat de exemple este bun, dar nu perfect. Uneori, modulul de detectare a nu reușește să detecteze mai mulți trecători în același timp.

Ca și direcții viitoare de dezvoltare, aplicației i se poate atașa un modul de urmărire a mișcărilor trecătorilor.

REFERINȚE

- [1] P. Viola and M. J. Jones, "Robust real-time face detection", in International Journal of Computer Vision, 2004.
- [2] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.
- [3] J. Leskovec. Detection of human bodies using computer analysis of a sequence of stereo images, Sentjost 1999.
- [4] S. Munder and D.M. Gavrila, An Experimental Study on Pedestrian Classification, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 28, NO. 11, NOVEMBER 2006
- [5] <http://cyberc3.sjtu.edu.cn/doc/paper/Robotica2006.pdf>
- [6] C. Papageorgiou and T. Poggio, "Trainable pedestrian detection", in Proceedings of the International Conference on ICIP, 1999
- [7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", in Proceeding of the IEEE International Conference on Computer Vision and Pattern Recognition, 2001.
- [8] <http://www.soe.ucsc.edu/classes/cmcs242/Winter08/proj/bweber-report.pdf>
- [9] <http://opencv.willowgarage.com/wiki/FaceDetection>
- [10] G. Monteiro, P. Peixoto, U. Nunes. Vision based pedestrian detection using haar-like features, in Robotica- Festival nacional de robotica, Guimaraes 2006