

Real-Time Human Detection Techniques

Oana Mihaela Vultur, Ștefan Gheorghe Pentiu

Abstract—Human detection is a challenging task with many applications that has attracted lot of attention in recent years [1]. Full body detection is used in application of video surveillance and security, where real-time systems are needed to analyse and process video sequences for intrusion detection. In this paper will try to present some techniques used for real-time human detection.

Index Terms—classification, computer vision, human tracking, pedestrian detection, urban traffic.

I. INTRODUCERE

DETECȚIA siluetei umane este o provocare, dar și o sarcină dificilă în același timp. Există multe aplicații care au atras atenția în ultimii ani și care realizează acest lucru. Una din aplicații ar fi supravegherea video și securizarea, aplicație în care este nevoie de sisteme care să funcționeze în timp real și care să analizeze și să proceseze secvențele video pentru a detecta eventualii intruși.

Un detector de obiecte poate fi văzut ca o combinație dintre două elemente cheie: un algoritm de extragere a caracteristicilor (care codifică regiuni de imagine sau părți din fișierele video ca și vectori de caracteristici) și un algoritm de detecție (care folosește caracteristicile calculate anterior pentru a decide dacă a detectat sau nu un obiect).

Principala dificultate în construirea unui detector performant de obiecte este nivelul de variație din imagine sau din fișierele video. Există câțiva factori care contribuie la acest lucru:

- Procesul de formare a imaginii reprimă din adâncimea 3D a imaginii și se crează o dependență de unghiul din care este privită imaginea, astfel încât chiar și cea mai mică modificare a poziției obiectului de interes sau a poziției camerei poate să schimbe considerabil imaginea.
- Cele mai multe clase de obiecte naturale au variații mari în interiorul clasei. De exemplu, în cazul siluetei umane, atât aspectul siluetei umane cât și postura se schimbă destul de mult între două imagini consecutive. Un detector performant trebuie să fie independent de aceste variații.
- Dezordinea din fundalul imaginilor este comună tuturor imaginilor și variază de la o imagine la alta. Ca exemplu avem acele scene în aer liber sau în mediile de interior. Detectorul folosit trebuie să fie capabil să distingă clasele de obiecte de fundalul încărcat.

- Culorile obiectelor precum și condițiile de iluminare variază considerabil de la o imagine la alta. De exemplu, lumina naturală, a soarelui și umbrele din timpul zilei față de lumina artificială a surselor de iluminat în timpul nopții. Astfel, un detector performant de obiecte trebuie să facă față schimbărilor de culoare și să nu fie influențat de schimbările de iluminare.
- Ocluziile parțiale creează dificultăți în procesul de detecție a obiectelor deoarece este vizibilă doar o parte a obiectului pentru a fi procesată.

În figura 1 sunt prezentate câteva exemple care ilustrează aceste dificultăți pe care le întâlnim în detecția siluetei umane. De remarcat este gama largă de posturi umane, variațiile în ceea ce privește aspectul siluetei umane și condițiile de iluminare, diferențele în ceea ce privește îmbrăcămintea și dezordinea de fundal.



Fig. 1. Imagini reprezentând oameni cu o gamă largă de variații în ceea ce privește aspectul, postura, îmbrăcămintea, iluminarea și fundalul. Unele imagini conțin și ocluzii parțiale[1].

II. DIVERSE TEHNICI DE DETECȚIE A SILUETEI UMANE

A. Sistemul de detecție a siluetei umane folosind caracteristici Haar și algoritmul de învățare AdaBoost, sistem propus de Paul Viola, Michael Jones și D. Snow

Unii dintre veteranii în domeniul recunoașterii formelor sunt Paul Viola și Michael Jones, care au realizat, în 2001, prima detecție în timp real a feței umane. Pentru a realiza acest lucru au folosit seturi de caracteristici de tip Haar și algoritmul de învățare AdaBoost. În 2003, Paul Viola, Michael Jones și D. Snow au propus un detector de siluete umane care conține descriptori de mișcare. Ei au construit un detector de persoane pentru aplicații de supraveghere video folosind camere statice, detector bazat pe unde Haar generalizate și același algoritm de învățare AdaBoost. Procesul de antrenare folosește AdaBoost pentru a selecta un subset de caracteristici și pentru a construi clasificatorul. La fiecare etapă algoritmul de învățare alege dintr-un set de filtre, incluzând și filtrele de înfățișare/aparență, filtrele de direcție a mișcării și filtrele de amplitudine a mișcării. Rezultatul algoritmului de învățare AdaBoost este un clasificator care constă într-o combinație liniară de caracteristici selectate. Un aspect important de reținut este acela că acest clasificator rezultat combină atât caracteristici de mișcare cât și caracteristici de aparență (înfațișare). La fiecare etapă AdaBoost alege din totalitatea caracteristicilor de mișcare și aparență, caracteristica cu cea mai mică eroare ponderată pe exemplele de antrenare. Clasificatorul rezultat echilibrează informațiile referitoare la intensitate și mișcare pentru a maximiza rata de detecție. Viola et al. propun o arhitectură de tip cascadă pentru detecția siluetei umane. Fiecare clasificator din cascadă este antrenat pentru a atinge rate de detecție ridicate. Detectorii mai simpli (care utilizează un număr mic de caracteristici) sunt poziționați în primele etape ale cascadei, în timp ce detectorii complecși (care utilizează un număr mare de caracteristici) sunt poziționați în etapele următoare ale cascadei. Detecția în cascadă vine de la simplu la complex.

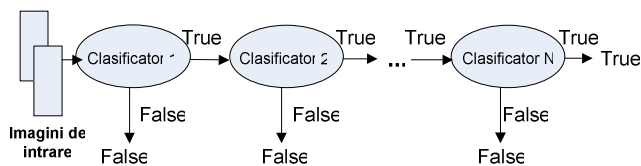


Fig. 2. Arhitectura cascadă folosită de Paul Viola, Michael Jones și D. Snow în cadrul sistemului de detecție a siluetei umane[2].

Fiecare etapă a cascadei constă dintr-un clasificator antrenat de algoritmul AdaBoost. Imaginile de intrare sunt trimise primului clasificator care decide „true” sau „false” (dacă este siluetă umană sau nu). O detecție falsă face ca detectorul să întoarcă valoarea „false” și imaginile respective nu mai sunt trimise mai departe următorului clasificator. O detecție adevărată are ca rezultat trimiterea imaginii de intrare la următorul clasificator din cascadă. Dacă toți clasificatorii returnează „true”, intrarea este clasificată ca o detecție. Dacă un singur clasificator din cascadă returnează „false”, intrarea este clasificată ca o detecție falsă. Arhitectura cascadă este foarte eficientă deoarece clasificatorii cu cele mai puține

caracteristici sunt poziționați la începutul cascadei, minimizând astfel numărul total de calcule.

Viola et al. au creat un set de secvențe video cu scene surprinse pe stradă, unde toți trecătorii sunt marcați de către un dreptunghi în fiecare frame. S-au folosit opt astfel de secvențe, fiecare conținând în jur de 2000 de frame-uri. Câteva frame-uri folosite pentru antrenare, precum și dreptunghiul galben care marchează manual silueta umană sunt prezentate în figurile 3, 4 și 5.



Fig. 3. Exemplu de frame folosit în procesul de antrenare [2].



Fig. 4. Un alt exemplu de frame folosit în procesul de antrenare [2].



Fig. 5. Exemplu de frame folosit în procesul de antrenare [2].

În setul de antrenare s-au folosit șase secvențe, iar din acest set a fost antrenat atât un detector dinamic de persoane, cât și

unul static. Alte două secvențe sunt utilizate pentru a testa acești detectori. Detectorul static de siluete umane folosește aceeași arhitectură de bază ca și cel de detecție a feței umane, descris în [3].

Rezultate ale detecției siluetelor umane folosind atât detectorul dinamic, cât și cel static sunt prezentate în figurile 6 respectiv 7.



Fig. 6. Exemple de detecții folosind detectorul dinamic. De remarcat ploaia și ninsoarea din imagine [2].



Fig. 7. Exemple de detecții folosind detectorul static [2].

B. Sistemul de detecție a trecătorilor propus de Ionela Rusu și Ștefan-Gheorghe Pentiu

Un alt sistem care realizează detecția posturii umane folosind caracteristici Haar și algoritmul de învățare AdaBoost este cel propus de Ionela Rusu și Ștefan-Gheorghe Pentiu (2010). 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 din fiecare rundă. Cel mai bun clasificator slab este clasificatorul care atinge cea mai mică eroare de formare[4]. În figura 8 este prezentată arhitectura sistemului de detecție a posturii umane.

Detecția 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 aceeași dimensiune, de posturi umane, numite imagini pozitive, cât și un număr de imagini ce nu conțin posturi umane, numite imagini

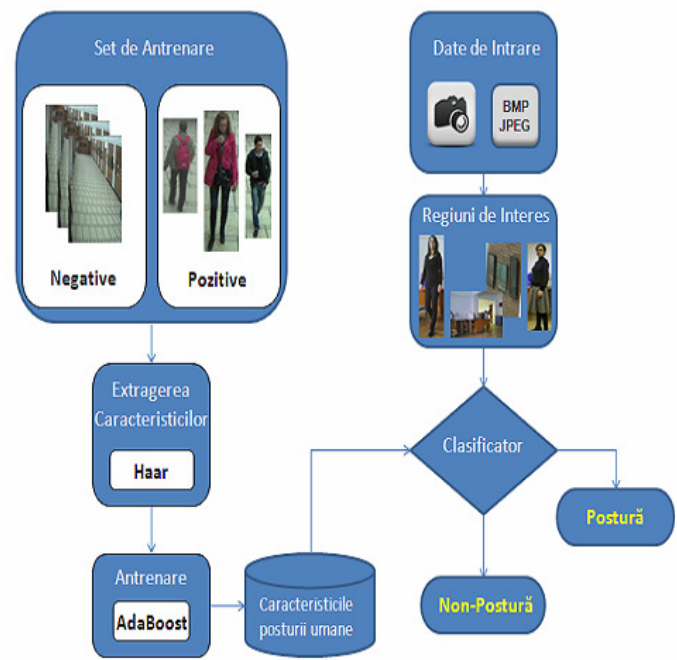


Fig. 8. Arhitectura sistemului de detecție a posturii umane [4].

negative[4]. În figura 9 este prezentat un subset de imagini pozitive, iar în figura 10 un subset de imagini negative din setul de antrenare folosit.



Fig. 9. Imagini pozitive folosite în setul de antrenare [4].



Fig. 10. Imagini negative folosite în setul de antrenare [4].

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 s-a folosit algoritmul AdaBoost. Acest algoritm va

stabilii caracteristica de tip Haar folosită [4]. Figura 11 prezintă un exemplu de detecție a posturii umane folosind sistemul mai sus menționat.

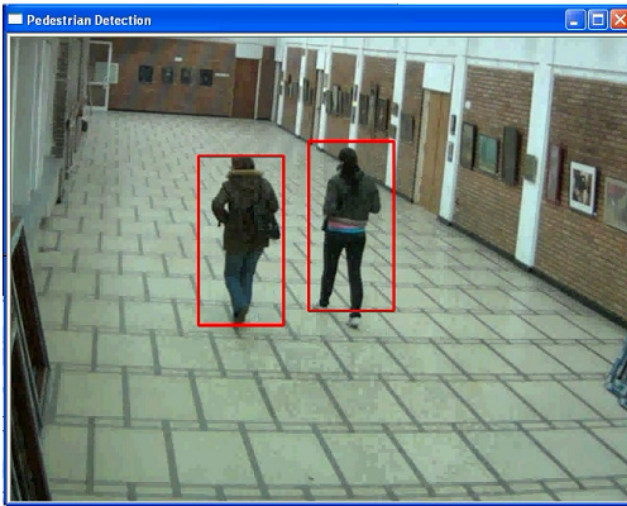


Fig. 11. Exemple de detecții folosind sistemul de recunoaștere al trecătorilor [4].

C. Sistemul de detecție a persoanelor SIPIVSP

SIPIVSP este unul dintre sistemele de detecție a siluetei umane care a fost dezvoltat în cadrul Universității „Ștefan cel Mare” – Suceava și este denumit pe larg: „Sistem inteligent de detectare a persoanelor și a traseului parcurs de acestea în înregistrările unui sistem de supraveghere video multi-cam”.

Sistemul SIPIVSP, după cum se poate observa și din figura 12, este alcătuit din cinci module:

1. modulul de achiziție a imaginilor;
2. modulul de detecție a persoanelor aflate în mișcare;
3. modulul de stocare a informațiilor referitoare la persoanele aflate în mișcare;
4. modulul web de interogare a bazei de date;
5. modulul de programare și declanșare a alertelor [5].

Acest sistem a fost dezvoltat în limbajul de programare Borland C++ Builder, iar pentru accesul la bazele de date s-a utilizat sistemul de gestiune al bazelor de date PostgreSQL. Testarea sistemului s-a făcut cu ajutorul a diferite înregistrări video preluate de la camerele unui sistem de supraveghere video instalat în capusul Universității „Ștefan cel Mare” din Suceava [5].

Sistemul SIPIVSP poate prelucra înregistrări video provenite de la diferite surse, cum ar fi: sisteme de supraveghere video, camere video de tipul webcam (conectate direct la computerul pe care rulează aplicația) sau un Hard-Disk. Detecția mișcării poate fi realizată atât pentru cazul imaginilor color, cât și pentru cel al imaginilor reprezentate pe 256 nivele de gri [5].

Modulul de detecție a persoanelor aflate în mișcare este responsabil cu realizarea următoarelor operații:

- detecția mișcării;
- detecția și eliminarea zgomotelor;
- identificarea persoanelor distincte;

- determinarea, pentru fiecare persoană aflată în mișcare, a valorilor unor parametri specifici (aria, coordonatele centrului de greutate etc);
- compresia imaginilor pentru stocarea acestora pe un server FTP (File Transfer Protocol) [5].

Operația de detecție a mișcării a fost realizată cu ajutorul algoritmilor de extragere a fundalului imaginilor. Detecția se realizează în două etape: învățare și recunoaștere [5].

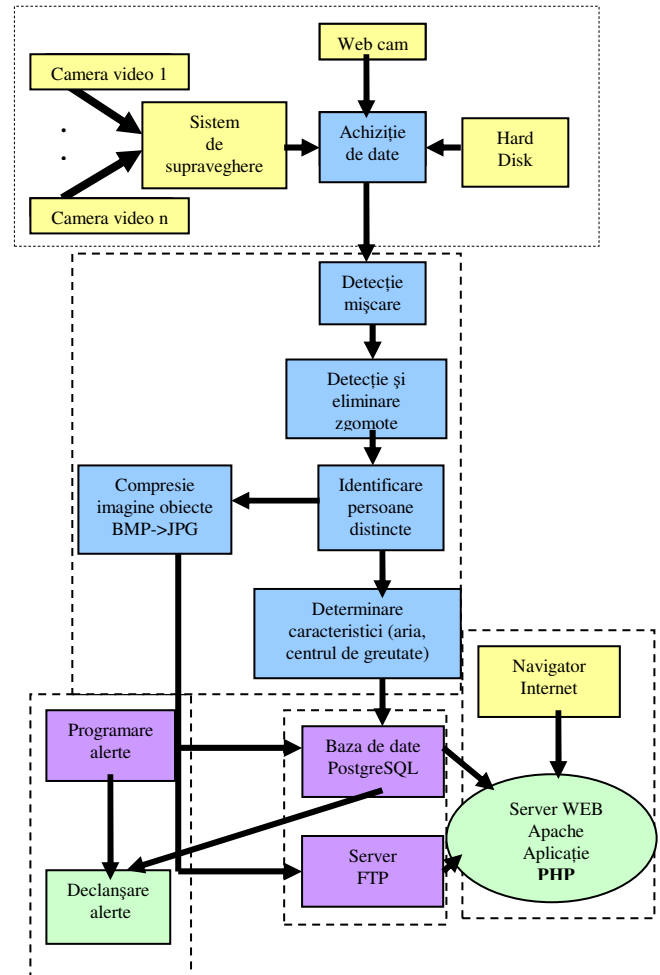


Fig. 12. Schema bloc a sistemului inteligent de detecție al persoanelor [5].



Fig. 13. Exemplu de detecție a persoanelor cu ajutorul sistemului SIPIVSP [5].

I. CONCLUZII

Sistemele de detecție a persoanelor prezentate în acest articol se bazează pe procesarea imaginii și tehnici computer vision. Ele pot detecta cu succes persoane din multiple unghiuri de vizualizare și au o rată de detecții false scăzută.

REFERINȚE

- [1] N. Dalal, "Finding People in Images and Videos", lucrare de disertație, Inst. Nat'l Polytechnique de Grenoble, 2006.
- [2] P. Viola, M. J. Jones și D. Snow, "Detecting pedestrians using patterns of motion and appearance.", in Proceedings of the 9th International Conference on Computer Vision, Nice, Franța, vol. I, pag. 734–741, 2003.
- [3] 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. W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
- [4] Ionela Rusu, Ștefan-Gheorghe Pentiuc, "Pedestrian Detection Using Vision-Based Techniques" , *Sisteme Distribuite (Suceava-online)*, p. 85-89, ISSN 2067 – 5259, Suceava, 2010.
- [5] Tudor Ioan Cerlincă, "Contribuții la dezvoltarea sistemelor inteligente de procesare a informațiilor video, spațiale și temporale", teză de doctorat, 2009.