IOSUD - Universitatea Politehnica Timişoara
Şcoala Doctorală de Studii Inginereşti

"CLASIFICAREA COMPORTAMENTULUI FOLOSIND REȚELE DE SENZORI VIDEO ÎN MEDIU URBAN"

Teză de doctorat – Rezumat

pentru obținerea titlului științific de doctor la
Universitatea Politehnica Timișoara
în domeniul de doctorat Calculatoare și Tehnologia Informației

autor ing. Marius BABA
conducător științific Prof.univ.dr.ing. Ionel JIAN

Cuprins

Scopul acestei lucrari este dezvoltarea unor algoritmi de analiză a secvențelor video pentru a permite detectarea automată a comportamentelor violente în mediu urban cu ajutorul unei rețele de senzori video. Algoritmii propuși în această lucrare utilizează tehnici de analiză a secvențelor video care nu necesită calcule complexe și astfel permit ca fiecare cameră din rețea să proceseze individual datele video colectate. Această abordare face posibilă analiza unui volum mare de date și totodată îmbunătățește rata de detecție a evenimentelor de interes de către sistemul de supraveghere.

1. Introducere

În acest capitol se prezintă motivația alegerii temei de cercetare. Cu toții suntem martori ai faptului că orașele devin din ce în ce mai aglomerate. Din acest motiv întâlnim pe străzi foarte frecvent conflicte între oameni. Cele mai numeroase dintre astfel de conflicte reprezintă comportamente agresive ale șoferilor din trafic și lupte de stradă între diverse grupuri infracționale. Pe lângă conflictele interumane, în mediul urban apar și o serie de alte evenimente violente precum atacuri ale câinilor vagabonzi asupra pietonilor.

Pentru a preveni astfel de situaţii specifice zonelor urbane se poate utiliza o reţea de senzori video împreună cu o metodă manuală pentru analiza secvențelor video. Ideea din spatele acestei abordări este că, odată detectat, evenimentul suspect este transmis în cel mai scurt timp către autoritățile competente pentru a se lua măsuri rapide în vederea combaterii acestor situații nedorite. Principala problemă a acestei abordări consta tocmai în analiza manuală a secvențelor video. Deobicei această sarcină este efectuată de către ofițerii de pază și ordine, care, cu ajutorul a mai multor monitoare conectate la diverse astfel de senzori video, inspecteaza vizual filmările capturate de senzori.

Inspectarea vizuală a secvențelor video nu este o metodă foarte eficientă. Ochii umani obosesc foarte repede, în special dacă își schimbă focalizarea atenției în mod constant. Din acest motiv foarte multe evenimente de interes rămân nedetectate. Conform experimentului prezentat în lucrarea [1] omul nu reușește să detecteze aproximativ 80% din acțiunile prezente într-o secvență video cu o durată de 30 de minute. Acest experiment este un indicator clar că procesarea manuală a secvențelor video nu este deloc eficientă. Mai mult, inspecția vizuală a unui volum mare de secvențe video în timp real este o operațiune costisitoare deoarece implică multă muncă umană.

O alternativă a acestei metode este analiza automată a secvențelor video cu ajutorul algoritmilor de tip *computer vision* (vedere artificială). În acest proces secvențele video sunt procesate de calculator și doar evenimentele de interes sunt transmise ofițerilor de pază. Această abordare permite detecția evenimentelor de interes cu mare precizie și face posibilă analiza în timp real a unui volum mare de date video la costuri mult mai mici.

Cu toate acestea, algoritmii de analiză a secvențelor video sunt foarte rar utilizați în sistemele de supraveghere video deoarece ridică o serie de probleme. Într-adevar, detecția evenimentelor presupune recunoașterea comportamentului uman în video ceea ce nu este o sarcină tocmai ușoară. Această operațiune implică tehnici complexe de analiză a secvențelor video care la rândul lor utilizează foarte multe resurse hardware. Din acest motiv majoritatea algoritmilor existenți în literatura de specialitate nu pot fi utilizați ca atare pentru analiza în timp real a unui volum foarte mare de secvențe video.

Mai mult, în multe cazuri comportamentele sunt ambigue, ceea ce împiedică rezolvarea cu mare acuratețe folosind astfel de algoritmi. Același set de acțiuni pot reprezenta comportamente total diferite. Spre exemplu acțiunea de alergare poate fi asociată cu un comportament normal în cazul în care individul analizat face sport sau poate fi asociată cu un

comportament anormal în cazul în care individul a comis un furt și aleargă pentru a nu fi prins de păgubaș.

Pentru a permite rețelelor de supraveghere să analizeze în timp real conținutul video, în această lucrare s-au propus trei algoritmi gândiți pentru analiza și detecția comportamentelor suspecte în zone urbane. Detectarea rapidă a evenimentelor periculoase este de o importanță deosebită deoarece face posibilă intervenția rapidă a ofițerilor de poliție, ceea ce înseamnă o îmbunătățire a securității publice în mediul urban.

2. Importanța recunoașterii comportamentelor umane

Capitolul 2 conține o scurtă descriere în ceea ce privește importanța recunoașteri comportamentelor. Tot aici sunt studiate taxonomiile folosite de algoritmii de detectare a comportamentelor umane în video și câteva sisteme de supraveghere care se folosesc de algoritmi de analiza a secvențelor video pentru a detecta acțiuni sau comportamente umane în acestea.

3. Structura tezei

Teza este organizată după cum urmează:

- Capitolul 1 prezintă o introducere în subiectul abordat în teză.
- Capitolul 2 discută importanța recunoașteri automate a comportamentelor umane în secvențe video de la sistemele de supraveghere urbană.
- Capitolul 3 detaliază structura tezei.
- Capitolul 4 prezintă stadiul actual în domeniul algoritmilor de analiza video, domeniul căruia i se adresează această teză.
- Capitolul 5 prezintă un prim algoritm de analiză a secvențelor video conceput în această lucrare. Algoritmul este dezvoltat pentru analiza traficului dintr-o intersecție aglomerată. Tot în acest capitol sunt prezentate și seturile de date video create cu scopul de a testa și valida eficiența algoritmului. Rezultatele obținute în urma rulării algoritmului sunt prezentate la sfârșitul capitolului.
- Capitolul 6 prezintă cel de-al doilea algoritm conceput în aceasta teză și care se folosește de proprietățile obiectelor în mișcare din video precum viteza, accelerația și traiectoria pentru a recunoaște atacul unor câini vagabonzi asupra unor persoane. La sfârșitul capitolului sunt de asemenea prezentate rezultatele obținute de algoritm.
- Capitolul 7 discută cel de-al treilea algoritm dezvoltat în această lucrare. Utilizând descriptori de mișcare extrași în procesul de decodare a secvențelor video encodate MPEG precum și metode de inteligență artificială bazate pe tehnici de învățare profundă, acest algoritm este capabil să recunoască comportamente umane complexe din medii urbane, cum ar fi de exemplu luptele de stradă. Rezultatele obținute de acest algoritm sunt expuse la sfârșitul capitolului.
- Capitolul 8 prezintă concluziile finale precum și lista de contribuții proprii.

4. Stadiul actual al domeniului

Capitolul patru este dedicat studiului algoritmilor existenți în literatura de specialitate. În acest capitol se studiază mai multe abordări de recunoaștere atât a acțiunilor cât și a comportamentelor umane în video.

Algoritmii studiați în acest capitol sunt împărțiți în două categorii:

- algoritmii proiectați manual
- algoritmii bazați pe tehnici de învățare profundă (deep learning)

Algoritmii proiectați manual se folosesc de tehnicile de analiză a secvențelor video și a imaginilor pentru a detecta acțiuni și comportamente umane în video. Acești algoritmi sunt compuși din câțiva pași auxiliari. Fiecare pas auxiliar analizează doar anumite aspecte ale acțiuni din video, iar rezultate analizelor sunt mai apoi combinate pentru a obține rezultatul final. Avantajul acestei metode este că permite proiectanților să aleagă tehnicile potrivite pentru a obține un algoritm performant de detecție a comportamentelor umane în video. În secțiunea dedicată acestor algoritmi, au fost studiați câțiva algoritmi reprezentativi de acest tip, precum algoritmi propuși în lucrările [2], [3] și [4].

În secțiunea dedicată algoritmilor bazați pe tehnici de învățare profundă, s-au prezentat concepte folosite de această tehnică precum și câțiva algoritmii de acest tip precum cei din [5], [6] și [7]. Algoritmii bazați pe învățarea profundă sunt compuși din o rețea neuronală artificială și o fază de post procesare. Rețeaua neuronală este destinată recunoașterii acțiunilor. Faza de post procesare, utilizînd un sistem expert bazat pe reguli, este destinată recunoașterii comportamentelor. Algoritmii bazați pe învățarea profundă nu fac uz de pași auxiliari. În schimb ei se folosesc de date de antrenare pentru a învăța caracteristicile entităților în procesul de clasificare a comportamentului uman în video.

5. Clasificarea a comportamentului de bază în video. Studiu de caz: Aplicație de supraveghere a traficului

În acest capitol este prezentat un algoritm proiectat manual pentru detectarea și urmărirea vehiculelor dintr-o intersecție aglomerată. Algoritmul conceput este publicat în articolul [8] și este un prim pas spre dezvoltarea algoritmilor avansați capabili să recunoască comportamentele umane complexe în filmările video.

La începutul cercetări s-au analizat scenarii simple precum cel privind circulația vehiculelor dintr-o intersecție aglomerată. Este cunoscut că direcțiile de circulație a vehiculelor sunt impuse de drumul pe care se circulă. Așadar, vehiculele se pot deplasa doar în anumite direcți și respectând anumite legi fizice, ceea ce implică comportamente in general simple (de bază). Într-adevăr, în acest scenariu sunt excluse traiectorii aleatoare, impredictibile.

Algoritmul propus în acest capitol este compus din următoarele părți funcționale:

- filtrarea zonelor
- extragerea obiectelor de prim plan folosind o mască
- corecția măștii de prim plan
- detecția de caracteristici ale vehiculelor
- faza de post procesare
- clasificarea și numărarea vehiculelor de fiecare tip

Partea algoritmului responsabilă pentru filtrarea zonelor are ca și scop eliminarea obiectelor irelevante din procesul de analiză. În această etapă se extrage porțiunea de drum pe care circulă vehiculele. Doar această porțiune este supusă procesului de analiză. Astfel, obiectele irelevante precum pietonii care circulă pe trotuar, blocurile din apropierea intersecției, sau copacii de lângă drum sunt ignorate de către algoritm.

Următorul pas este extragerea obiectelor de prim plan, care, așa cum sugerează și numele, vizează separarea obiectelor de prim plan de obiectele de fundal. Această sarcină este realizată de algoritmul de tip MOG (Mixture of Gaussians) propus în articolul [9]. Este un algoritm foarte frecvent utilizat în tehnicile de procesare a secvențelor video deoarece obține rezultate excelente. Actualizarea periodică a modelului de fundal îi permite să se adapteze foarte ușor la condițiile de iluminare schimbătoare ale scenei analizate. Astfel, obiectele de prim plan sunt separate cu mare precizie de obiectele de fundal.

În ciuda preciziei sale ridicate, algoritmul MOG nu reușește întotdeauna să clasifice corect toți pixelii. Foarte frecvent, masca de prim plan generată de algoritmul MOG conține și un număr mic de pixeli clasificați greșit care afectează negativ procesul de detecție a vehiculelor. Pentru a remedia această problemă, masca de prim plan este corectată cu ajutorul operatorilor morfologici. Operatorii morfologici sunt foarte utili deoarece nu necesită calcule complexe și reusesc să reducă semnificativ numărul de pixeli clasificați greșit.

Următoarea etapă este responsabilă pentru extragerea de caracteristici ale vehiculelor. În această fază algoritmul analizează obiectele de prim plan și asociază fiecărui obiect detectat un dreptunghi de încadrare. Dreptunghiurile de încadrare sunt calculate utilizând contururile obiectelor care sunt extrase din masca de prim plan folosind algoritmul propus în articolul [10].

Caracteristicile vehiculelor sunt mai apoi filtrate. Operațiunea respectivă este efectuată în faza de post procesare și are ca și scop eliminarea obiectelor ce nu reprezintă vehicule din procesul de analiză al algoritmului. Tot în cadrul acestei faze se regăsește și metoda de corectare a contururilor extrase greșit.

Ultima parte a algoritmului este responsabilă pentru clasificarea tipului vehiculelor și apoi numărarea acestora. În această fază vehiculele sunt clasificate fie în vehicule mari fie în vehicule mici și mai apoi numărate. Vehicule mici reprezintă clasa autoturismelor, în timp ce vehicule mari reprezintă clasa vehiculelor comerciale de mari diemnsiuni (camioanelor).

Rezultatele obținute de algoritm precum și detaliile aferente celor două seturi de date video create în această teză sunt prezentate la sfârșitul acestui capitol.

6. Clasificarea comportamentelor în video utilizând mișcarea obiectelor și proprietăți ale acesteia

Capitolul 6 descrie cel de-al doilea algoritm conceput în această teză care a fost publicat în articolul [11]. Acesta este dezvoltat pentru recunoașterea comportamentelor complexe în video precum atacurile câinilor vagabonzi asupra oamenilor.

Similar cu algoritmul introdus în capitolul anterior, algoritmul propus în acest capitol este împărțit la rîndul său în mai multe părți funcționale. Operațiunile au fost grupate după tip rezultând astfel:

- grupa unu- responsabilă pentru procesările de nivel scăzut
    - extragerea obiectelor de prim plan
    - clasificarea obiectelor

- grupa doi - menită pentru procesările de nivel înalt
    - extragerea traiectoriilor și a caracteristicilor mișcării
    - detectarea evenimentul de atac

Operațiunea de extragere a obiectelor de prim plan este efectuată folosind tot algoritmul MOG. Masca generată de acest algoritm este mai apoi corectată utilizând operatori morfologici. Această tehnică a fost preluată din algoritmul prezentat în capitolul anterior deoarece a generat rezultate bune.

Obiectele de prim plan sunt mai apoi clasificate. În această etapă, contururile obiectelor sunt extrase cu ajutorul algoritmului descris în articolul [10] și apoi clasificate. Clasificarea este efectuată folosind o tehnică robustă, dezvoltată în această lucrare, care nu folosește calcule complexe și oferă o clasificare satisfacătoare a obiectelor.

Extragerea traiectorilor și a proprietăților sale este următoarea fază a algoritmului. În această etapă, pe lângă extragerea traiectorilor, algoritmul extrage vitezele și accelerațile obiectelor. Conform analizei efectuate în acestă teză, traiectorile nu sunt suficient de descriptive pentru recunoașterea comportamentelor complexe. Din acest motiv pe lângă traiectorii s-au folosit caracteristici adiționale precum viteza și accelerația obiectelor.

Astfel, pentru fiecare punct al traiectoriei unui obiect, algoritmul generează un vector compus care conține poziția, viteza și accelerația obiectului în acel moment dat. Vectorii rezultați sunt mai apoi furnizați unui clasificator de tip SVM (support vector machines) care este responsabil pentru detectarea evenimentului de atac.

Ultima secțiune a acestui capitol conține descrierea seturilor de date utilizate în experimente și discută rezultatele obținute de algoritm.

7.Clasificarea comportamentelor complexe utilizând tehnica de învățare profundă și caracteristici proiectate manual

În vederea recunoașteri comportamentelor complexe în video, a fost conceput și un al treilea algoritm care este descris în detaliu în acest capitol. Algoritmul propus, publicat în articolul [12] este unul hibrid și aduce îmbunătățiri evidente celui din capitolul anterior. Acesta combină metodele folosite de algoritmii proiectați manual cu metodele utilizate de algoritmii bazați pe învățarea profundă. Abordarea respectivă permite algoritmului să învețe rapid clasificarea cadrelor video și face ca procesul de inferență să nu fie pretențios în ceea ce privește utilizarea resurselor hardware.

Procesul de analiză a algoritmului constă din următoarele etape:

- extragerea caracteristicilor de mișcare
- clasificarea cadrelor video
- filtrarea rezultatelor

Etapa întâi a algoritmului folosește codecul video MPEG pentru a extrage caracteristicile de mișcare. În procesul de decodare, codecul furnizează un set de vectori de mișcare care captează cu succes mișcările obiectelor în video. Această proprietate a codecului este benefică deoarece oferă informații despre mișcarea obiectelor pe în mod cvasi-gratuit (nu necesită procesări complexe). Caracteristicile de mișcare extrase în acest fel poartă denumirea de flux MPEG (MPEG flow).

Următoarea etapă este dedicată clasificării cadrelor video. Pentru această operațiune algoritmul folosește o rețea convoluțională profundă. Rețelele de acest tip sunt capabile să învețe caracteristicile și procesul de clasificare și astfel reușesc să clasifice cu mare precizie cadrele video. În ciuda acestei proprietăți benefice, rețelele neuronale au și ele un dezavantaj. Pentru buna funcționare, rețeaua trebuie instruită cu un volum foarte mare de date. Acest proces este unul laborios deoarece implică foarte multă muncă. Cu toate acestea, procesul de învățare poate fi simplificat prin folosirea caracteristicilor de mișcare extrase manual. Astfel, pentru a reduce volumul de date necesare pentru instruire, rețeaua utilizată de algoritm folosește caracteristici de mișcare extrase manual, aici sub formă de descriptori de mișcare tip MPEG flow.

O acțiune de luptă este compusă de obicei din mișcări violente urmate de momente de pauză. Din acest motiv, predicțiile rețelei oscilează în condițiile de prezență a luptei în video. Acest comportament al algoritmului nu este tocmai potrivit deoarece de la un astfel de algoritm se așteaptă să furnizeze predicții stabile. Pentru a remedia această problemă s-a conceput filtrul în domeniul timp care reușește cu succes să stabilizeze predicțiile rețelei.

Pe lângă descrierea algoritmului, capitolul 7 conține și o secțiune dedicată comparației descriptorului MPEG flow cu descriptorul de mișcare generat de fluxul optic care este de altfel foarte frecvent utilizat pentru detecția comportamentelor umane în video. Ultima porțiune a

acestui capitol conține descrierea seturilor de date utilizate în experimente, procedura de antrenare a rețelei și rezultatele obținute de algoritm.

8. Concluzii

Capitolul 8 conține concluziile lucrări. Tot aici este și o secțiune în care sunt prezentate contribuțiile personale, precum:

- S-a propus o metodă de detectare a comportamentului violent în zonele urbane care utilizează algoritmi de învățare profundă și o rețea de senzori video cu resurse de calcul limitate.
- S-au studiat diferiți algoritmi pentru urmărirea entităților în mișcare în video, urmărind performanța acestora relativ la problema studiată. Studiu de caz implementat: aplicație de supraveghere a traficului urban.
- S-a investigat detectarea comportamentelor violente în mediul urban. Studiu de caz implementat: atacul câinilor vagabonzi asupra unei persoane.
- S-a comparat performanța descriptorului de mișcare generat de fluxul optic cu perfomanța descriptorul tip MPEG flow.
- S-a propus o metoda de augmentare pentru mărirea setului de date video.
- S-a propus o metodă pentru clasificarea și numărarea vehiculelor în video independent de comportamentul șoferilor.
- S-a propus o metodă de corecție pentru rezolvarea problemei de împărțire a conturului obiectului. Această problemă este frecvent întâlnită la algoritmii de supraveghere a traficului și afectează în mod negativ performanța acestora.
- S-a propus o metodă pentru extragere și clasificare a entităților (ex. oameni vs.câini) în video.
- S-au comparat diferiți algoritmi pentru extragerea obiectelor de prim plan.
- S-au creat două baze de date video utilizabile pentru testarea aplicațiilor de supraveghere a traficului.

BIBLIOGRAFIE SELECTIVĂ

[1] D. Elliott, “Intelligent video solution: a definition”, Security Magazine, 47(6), pp.46–48, June 2010.

[2] Serhan Cosar, Giuseppe Donatiello, Vania Bogorny, Carolina Garate, Luis Alvares, François Bremond. "Toward abnormal trajectory and event detection in video surveillance." IEEE Transactions on Circuits and Systems for Video Technology 27.3 (2016): 683-695.

[3] Gorelick, Lena, et al. "Actions as space-time shapes." IEEE transactions on pattern analysis and machine intelligence 29.12 (2007): 2247-2253.

[4] Huangkai Cai, He Jiang, Xiaolin Huang, Jie Yang, “Violence Detection based on Spatio-Temporal Feature and Fisher Vector”, Chinese Conference on Pattern Recognition and Computer Vision (PRCV) 2018.

[5] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In Advances in Neural Information Processing Systems, pages 568–576, 2014.

[6] Yue-Hei Ng, Joe, et al. "Beyond short snippets: Deep networks for video classification." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[7] Zhang, Bowen, et al. "Real-time action recognition with deeply transferred motion vector cnns." IEEE Transactions on Image Processing 27.5 (2018).

[8] Áron Virginás-Tar, **Marius Baba**, Vasile Gui, Dan Pescaru, Ionel Jian, “Vehicle Counting and Classification for Traffic Surveillance using Wireless Video Sensor Networks”, Conference: 22nd Telecommunications Forum (TELFOR 2014), 25-27 November 2014, SAVA Center, Belgrade, Serbia; Published in IEEE XPLORE Digital Library; Indexed in the ISI Web of Science.

[9] Z. Zivkovic. “Improved adaptive Gaussian mixture model for background subtraction”, In the proceedings of the 17th International Conference on Pattern Recognition ICPR’04, 2004.

[10] Suzuki, Satoshi. "Topological structural analysis of digitized binary images by border following." Computer Vision, Graphics, and Image Processing 30.1 (1985): 32-46.

[11]**Marius Baba**, Dan Pescaru, Vasile Gui, Ionel Jian, “Stray Dogs Behaviour Detection in Urban Area Video Surveillance Streams”, Conference: 12th IEEE International Symposium on Electronics and Telecommunications (ISETC 2016), 27-28 October 2016, Timisoara, Romania; Published in IEEE XPLORE Digital Library; Indexed in the ISI Web of Science.

[12]**Marius Baba**, Vasile Gui, Cosmin Cernazanu, Dan Pescaru “A Sensor Network Approach for Violence Detection in Smart Cities Using Deep Learning”, Journal: Sensors, Volume 19, Issue 7; Published by MDPI Switzerland, 8 April 2019. ISI journal Q1 (Instruments and instrumentation). Journal impact factor 3.735.