

KÉPFELDOLGOZÁSI MÓDSZEREK ALKALMAZÁSA SZEMÉLY-SZÁMLÁLÁSI FELADATOKBAN

APPLICATION OF IMAGE-PROCESSING METHODS IN PERSON-COUNTING

Kovács Tamás^{ORCID:0000-0003-3947-8696} ^{1*}, Agg Péter^{ORCID:0000-0003-0148-2868} ¹,
Bolla Kálmán^{ORCID:0000-0002-4666-0990} ¹, Csizmás Edit^{ORCID:0000-0003-4397-1758} ¹,
Kovács Márk^{ORCID:0000-0002-9482-7695} ¹, Medgyes Krisztián^{ORCID:0000-0001-9716-2724} ¹,
Pásztor Attila^{ORCID:0000-0001-7354-5114} ¹

¹ Informatika Tanszék, GAMF Műszaki és Informatikai Kar, Neumann János Egyetem, Magyarország
<https://doi.org/10.47833/2022.2.CSC.005>

Kulcsszavak:

mesterséges intelligencia;
képfeldolgozás;
vizuális objektum azonosítás

Keywords:

artificial intelligence, image
processing,
visual object identification

Cikktörténet:

Beérkezett 2022. október 10.
Átdolgozva 2022. október 31.
Elfogadva 2022. november 5.

Összefoglalás

Ebben a munkában ipari célú személy-számlálási feladatok képfeldolgozáson alapuló megoldási lehetőségeit mutatjuk be. Az itt bemutatott módszerek elsősorban beltérben, illetve nem túl gyorsan változó vizuális információ esetén alkalmazhatóak, és három fő megoldási módszerre koncentrálnak. Ezek a módszerek: a mozgásdetekció, a háttér-szeperáció illetve a „deep-learning” alapú objektum felismerés. A munka során felmérésre került ezen módszerek megbízhatósága, és alkalmazhatósági köre, az adott módszer erősségeivel és gyengeségeivel együtt.

Abstract

In the present work a detailed survey of the visual methods for identification of people presence and visual people counting is given. The main focus of the paper is on the application of the motion-detection, the background-segmentation and the deep-learning object detection algorithms.

1. Bevezetés

Az emberi jelenlét érzékelése illetve a jelenlévő személyek számának becslése gyakran előforduló feladat az iparban vagy egyéb irodai felhasználási területeken. Ez két független feladatot jelent. Egy ilyen feladat gyakran felmerül olyan automatizált terekben, helyiségekben, ahol az automatikus rendszer nagyobb fogyasztású elemei csak akkor aktivizálódnak, ha ember kerül a közelébe. A személyek számának becslése is felmerül feladatként például a tömegközlekedésben a járatra várakozó vagy a járműben utazó utaslétszám becslésénél. Emellett, és ez lesz a jelen kutatási beszámoló fókuszterülete, fontos lehet a jelenlévő személy-létszám becslése intelligens épületgépészeti alkalmazások fejlesztéséhez. Ez a kutatási beszámoló a Neumann János Egyetem és az Airvent Légtechnikai Zrt. Közös fejlesztési eredményeit foglalja össze a képfeldolgozási megoldások területén. Ennél a problémánál a létszám becslése a légfertőtlenítő berendezésnek ad majd visszacsatolást a teljesítmény optimális megválasztása céljából.

* Kapcsolattartó szerző.
E-mail cím: kovacs.tamas@gamf.uni-neumann.hu

Közel húsz évvel ezelőtt publikált Kim és munkatársai [1] egy személy-azonosító és követő rendszert, mely egy biztonsági kapura rögzített kamera képét dolgozta fel, az akkori technikai háttér mellett kb. 10 fps (frames per second) sebességgel, ami valós idejűnek tekinthető. A cél a kapu előterében áthaladó emberek számlálása irány szerint is elkülönítve.

Ez az alkalmazott séma azóta sok hasonló, nem „deep learning” alapú megoldásnak lett a kiinduló pontja. Hsieh és munkatársai [2] is ezt a sémát alkalmazták több javítással: a háttérbecslést egy statisztikai módszerrel végezték (Gaussian Mixture Pixel Probability, kevert Gauss eloszlás módszere) így az jóval robusztusabb lett. Az előtér meghatározásnál pedig egy hiba-javító algoritmust alkalmaztak, amely kiszűri a kamera kisebb rezgéseiből adódó zavart. Emellett a háttérkivonás után a háttérkép változó árnyékait is megbecsülték és eliminálták. Az ezt követő mintázat illesztés alapú objektum azonosításnál is több gyorsító módosítást vezettek be, amely sokat segített a valós idejű alkalmazásoknál.

Egy 2010-ben közzétett munkájában Hou és Grantham [3] ugyanúgy a kevert Gauss eloszlás módszerét alkalmazta a háttér kivonásnál. Ebben a munkában az azonosítandó objektumok a képen meglehetősen kis méretű ember-alakok voltak (kb. 10-15 pixel átmérőjű objektumok), azaz az objektum azonosítás jelentette a legnagyobb kihívást. Hou és Grantham a Kanade-Lucas-Tomasi (KLT) sarokpont keresési eljárást alkalmazásával jelölte ki az alakok jellegzetes pontjait, majd a klaszterezési eljárással rendelte őket egy objektumhoz. Ezen klaszterek geometriai jellemzői alapján azonosította a humán alakokat a képen (többek közt a burkoló ellipszis kis- és nagytenyely méretei alapján).

2017-ben Anshari és Shim [4] a fenti feldolgozási sémát alkalmazta, de a háttérkivonás és bináris transzformálás után egyszerű folt detektáló eljárással azonosították a humán jelenlétet, és ezzel az egyszerűsítéssel alkalmassá tették az algoritmust arra, hogy olyan kisebb teljesítményű beágyazott hardveren is működjön, mint a Raspberry Pi 3.

Ezek mellett az egyszerűbb heurisztikus megoldások mellett kidolgozásra kerültek betanítást igénylő (deep learning) osztályozási módszerek is. Ilyen például Al-Zaydi és munkatársai megoldása [5], ahol egy tanítható regressziós modell alapján állapítják meg a képen látható személyek számát, amelynél a tanítást a Gauss Process Regression eljárás alapján végezték. Wahyuni és munkatársai [6] pedig szintén egy tanítható Support Vector Machine osztályozót használtak az objektum azonosításhoz.

Kanatov [7] és egy évvel később Nogueira [8] a manapság igen elterjedten alkalmazott konvolúciós neurális háló (CNN) alapján történő objektum felismerési technikát használták, amely a betanítási folyamatban nagy számítási kapacitást igényel ugyan, de az ismert CNN együtthatók alapján a rendszer implementálható egyszerűbb hardver eszközökre is.

A továbbiakban a feladat megoldására szóba jövő képfeldolgozási módszereket mutatjuk be a általunk implementált rendszerek teszt-eredményeivel együtt.

2. A feladat részletei

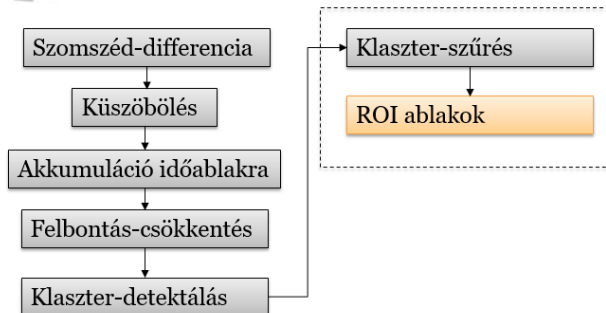
Ahhoz, hogy a légfertőtlenítő berendezés megfelelő információt kaphasson, valós idejű, és 1 frame per másodpercnél nem ritkább robosztus mérést kell megvalósítani. A használható szenzor: egy darab fixen felszerelt és fix tengelyű RGB kamera, széles látószöggel, amely lehetőleg a falon elhelyezett berendezésen van telepítve. A kamera és a kapcsolódó egyéb hardver elemek költsége a 100 EUR nagyságrendet nem haladhatja meg.

A megfigyelt helyiségek méretét illetően széles skálán kell megoldásokat találni: a néhány fős irodai helyiségtől a 10-20 fős tárgyalóig vagy tanteremig. Ennél nagyobb terekben összegyűlt tömeg létszám-becslése egészen más módszereket igényel. A célkitűzés megfogalmazásában a megoldások pontosságát tekintve 10-20%-os mérési hibát még elfogadhatónak tekintünk. Ebben a cikkben a pontosság szisztematikus mérésével nem foglalkozunk, mert ehhez egy nagyságrenddel nagyobb számú tesztre lenne szükség, mint ami itt bemutatásra kerül. A bemutatott tesztek eredményei csupán előzetes információt szolgáltatnak a megoldások használhatóságát és a fejlesztési irány helyességét illetően.

A továbbiakban bemutatott tesztekben a kis helyiséghez egy mobil-telefon kamerával rögzített 720p videókat használtunk, a nagy teremhez pedig egy 16 mm-es lencse átmérőjű 90 fokos látószögű kamerával készült, szintén 720p felbontású videókat.

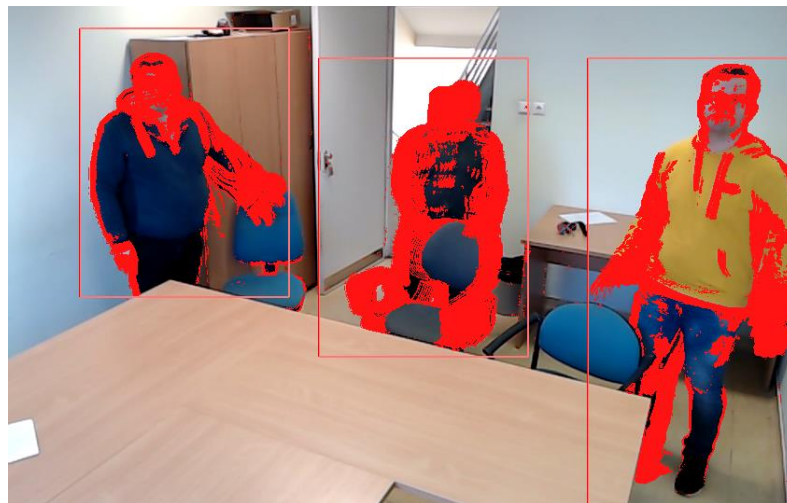
3. Integrált frame-differencia módszer

Ennek a megoldásnak az algoritmusát az 1. ábra szemlélteti. A közvetlen egymást követő képek (frame-ek) differenciájának abszolút értékét felül-áteresztő küszöb szűrés után összegezzük egy fix időablakban (melynek hossza néhány másodperctől akár 10 mp-ig terjedhet). Ez egy gyorsan futó algoritmus, sebessége egy átlagos PC-n (pl.: 8th gen. Intel Core I7 CPU, GPU alkalmazás nélkül) is eléri a 30 fps értéket.



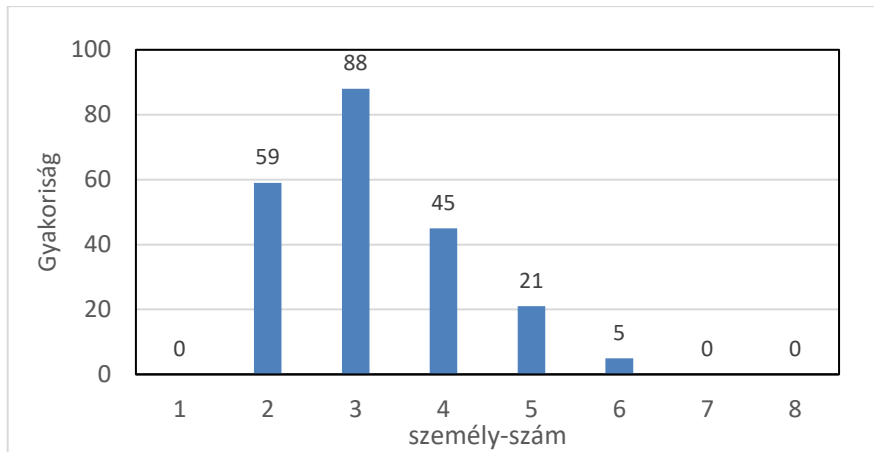
1. ábra: Az integrált frame-differencia módszer algoritmusának blokk ábrája

Ez a módszer a videó által rögzített mozgásra érzékeny, tehát egy vizuális mozgásdetekciós módszernek tekinthető. A fenti leírás alapján kapott mozgásdetekciós képeken a felbontás csökkentése (kb. 90x60 vagy kisebb) után a jelentősebb klasztereket kell megkeresni és alak, valamint sűrűség szerint szűrni, majd a klaszterek alapján a mozgó objektumok száma megállapítható (lásd 2. ábra).



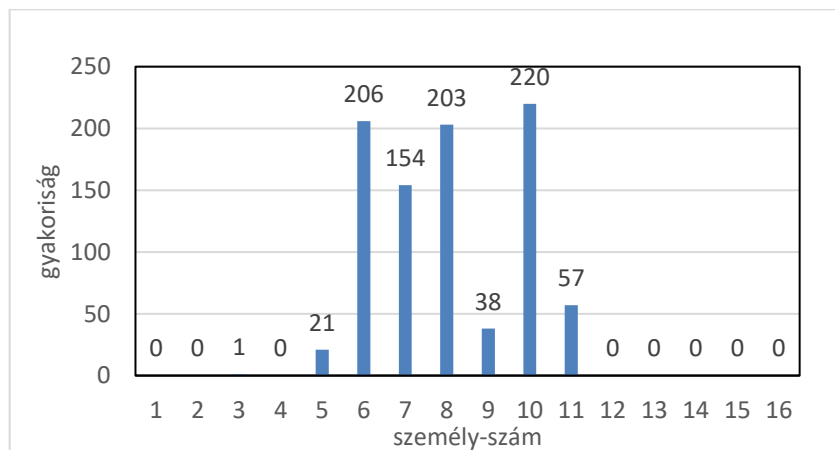
2. ábra: Az integrált frame-differencia módszer demonstrációja egy kis méretű iroda esetén, az alkalmazott időablak hossza 2 mp

Kisebb helyiségekben a módszer önmagában alkalmazva is viszonylag hatékonynak mondható. Egy kis helyiségben 3 főről készült felvételen teszteltük a módszert, és minden frame-en meghatároztuk a detektált személy-számot és egy 15 másodperces mérésben, majd ábrázoltuk a kapott személy-számok gyakoriságát (3. ábra). Látható, hogy a maximum 3, azaz ebben a környezetben működőképes ez az algoritmus



3. ábra: A teszt-mérés során detektált személy-szám gyakoriságok kis teremben 3 fő jelenléte esetén

Nagyobb terem és 10-15 fős létszám esetén azonban jelentősen romlik a bemutatott módszer hatékonysága, ami elsősorban a pixel felbontás csökkenésének köszönhető. A személyek alacsony pixel-felbontása mellett ugyanis a kisebb mozgások nem jelennek meg a képen. A 4. ábrán a fentebb leírt módszerrel elvégzett 60 másodperces teszt eredményét láthatjuk egy kb. 50 nm-es számítógépes tanterem és 14 fő esetére.



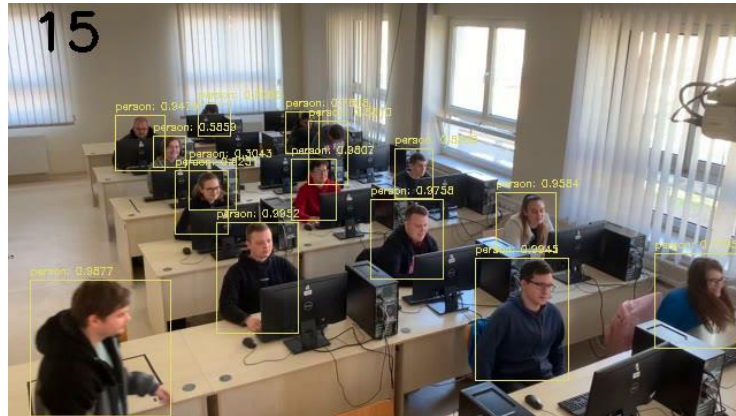
4. ábra: Nagy terem és 14 fős teszt-mérés során detektált személy-szám gyakoriságok a mozgásdetekciós módszerrel mérve

Látható, hogy a mért személy-szám gyakoriságok nem mutatnak egy meghatározó maximumot, hanem több személy-szám esetén is magas értékeket kapunk, és ezek a 6-11 személy-szám tartományba esnek, azaz jelentősen alacsonyabbak a várt 14 fős értéknél.

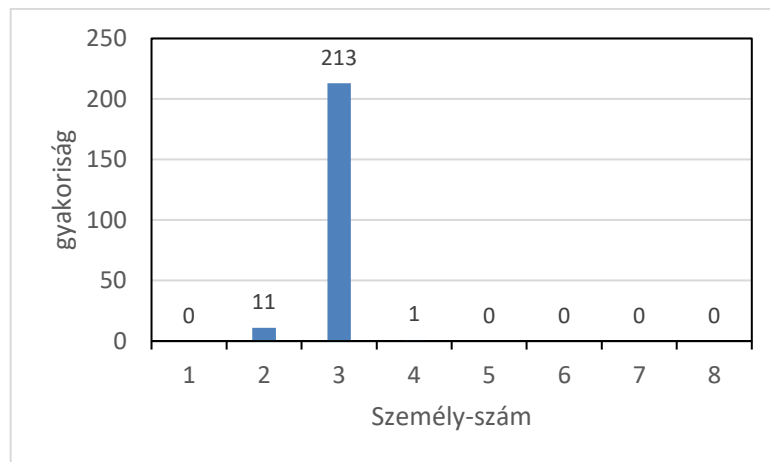
4. Deep-learning alapú személy-számlálás

Ennél a módszernél egy betanított konvolúciós elvű neurális hálót (CNN) használunk a személy-objektumok felismeréséhez a kiválasztott videó frame-eken. Maga a hálózat könnyen felépíthető a Pytorch vagy a Tensorflow Python könyvtárak alapján, a betanító kép-halmaz elkészítése és maga a betanítás folyamata azonban igen munka és időigényes folyamat. Ezért használtunk egy már betanított hálózatot, amely ingyenesen elérhető a git-hub-on [9, 10].

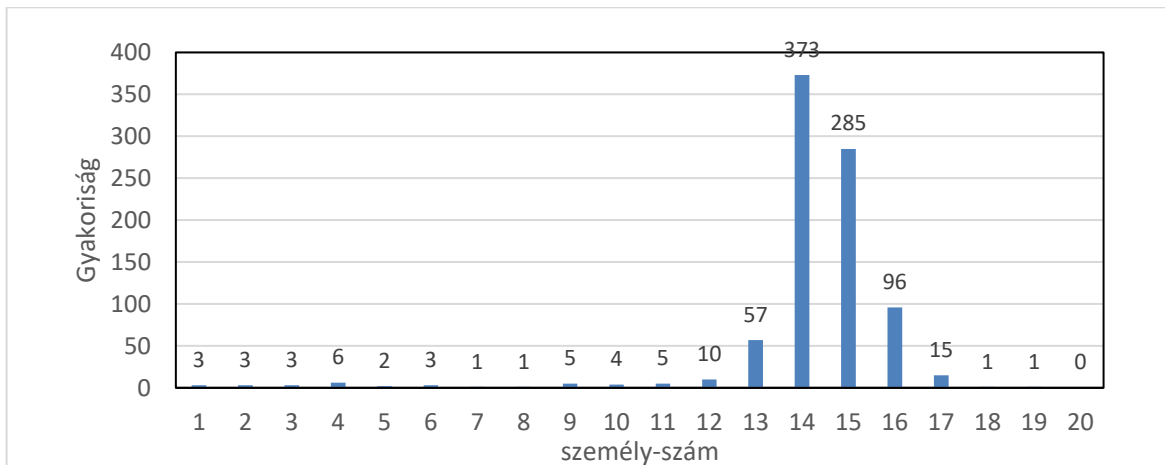
A fentebb is bemutatott 3 és 14 fős tantermi tesztet ezzel a módszerrel is elvégeztük és az eredmények jelentősen jobbak lettek a mozgásdetekciós módszerhez viszonyítva (5., 6. és 7. ábra): az egyértelmű maximum-helyek ezúttal a helyes 3 és 14-es értéknél vannak és a 14 fős esetben ezen a maximumon kívül csak egy ilyen nagyságrendű érték van a 15-ös személy-számnál.



5. ábra: A deep-learning alapú objektum felismerő módszer demonstrációja egy tanterem esetén 14 fő jelenlétével.



6. ábra: Kis terem és 3 fős teszt-mérés során detektált személy-szám gyakoriságok a deep-learning módszerrel mérve



7. ábra: Nagy terem és 14 fős teszt-mérés során detektált személy-szám gyakoriságok a deep-learning módszerrel mérve

5. Konklúzió

A kutatási munka során az itt bemutatott módszereken kívül implementáltuk és teszteltük még a nem deep-learning alapú alakzat-felismerési módszereket (HOG és HAAR alapú algoritmusok),

továbbá a háttér-előtér szegmentáláson alapuló áthaladási számlálási módszert. Az eredmények alapján az HOG és HAAR alapú módszerek önállóan nem, de az előtér-háttér szétválasztáson alapuló akár önálló módszerként is alkalmazható a feladat megoldására. Ezen eredmények részletesebb bemutatására ebben a publikációban már nincs lehetőség.

A részletesen bemutatott (mozgás-detekciós és deep-learning objektum felismerő) módszerek közül mindkettő alkalmas a feladatra, de a mozgásdetekciós módszer használhatósága elsősorban kisebb, néhány fős helyiségekre korlátozódik. Előnye viszont a nagy képfeldolgozási sebesség és a kisebb mikrovezérlő board-okra való alkalmazhatóság. A deep-learning módszer használhatósága kiterjed akár a 15-20 fős helyiségekig is, hátránya viszont, hogy nagy számítási igényű, így csak nagyobb teljesítményű, drágább hardver eszközön futtatható.

Köszönetnyilvánítás

A szerzők köszönetet mondanak a projektben résztvevő intézmények - AIRVENT ZRT és a Neumann János Egyetem GAMF Műszaki és Informatikai Kar - kollégáinak. Köszönettel tartozunk a kutatás támogatásáért, amely a " Széles körben használható levegő sterilizáló megoldások kifejlesztése intelligens működés optimalizáló vezérléssel 2020-1.1.2-PIACI-KFI-2021-00294 " pályázat keretében valósult meg. A projekt a Magyar Állam és az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával, a Széchenyi 2020 program keretében valósul meg.

Irodalomjegyzék

- [1] Kim, Jae-Won, et al. "Real-time vision-based people counting system for the security door." Proceedings of the IEEK Conference. The Institute of Electronics and Information Engineers, 2002.
- [2] Hsieh, Jun-Wei, Cheng-Shuang Peng, and Kao-Chin Fan. "Grid-based template matching for people counting." 2007 IEEE 9th Workshop on Multimedia Signal Processing. IEEE, 2007. DOI: 10.1109/MMSP.2007.4412881
- [3] Hou, Ya-Li, and Grantham KH Pang. "People counting and human detection in a challenging situation." IEEE Transactions on Systems, Man, and Cybernetics-part A: Systems and Humans 41.1 (2010): 24-33. DOI: 10.1109/TSMCA.2010.2064299
- [4] Ansari, Md Israfil, and Jaechang Shim. "People Counting System using Raspberry Pi." Journal of Multimedia Information System 4.4 (2017): 239-242.
- [5] Al-Zaydi, Zeyad, Branislav Vuksanovic, and Imad Habeeb. "Image processing based ambient context-aware people detection and counting." International Journal of Machine Learning and Computing 8.3 (2018): 268-273. DOI: 10.18178/ijmlc.2018.8.3.698
- [6] Wahyuni, Elvira Sukma, Rizqi Renafasih Alinra, and Hendra Setiawan. "People counting for indoor monitoring." 2017 International Conference on Computing, Engineering, and Design (ICCED). IEEE, 2017. DOI: 10.1109/CED.2017.8308112
- [7] Kanatov, Maksat, and Lyazzat Atymtayeva. "Deep convolutional neural network based person detection and people counting system." Advanced Engineering Technology and Application 7.3 (2018): 9-16. DOI 10.21608/aeta.2018.200330
- [8] Nogueira, Valério, et al. "RetailNet: A deep learning approach for people counting and hot spots detection in retail stores." 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). IEEE, 2019. DOI: 10.1109/SIBGRAPI.2019.00029
- [9] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018). <https://github.com/ultralytics/yolov5>. DOI 10.48550/arXiv.1804.02767