Huszár Viktor Dénes[ℓ]

Hamisítás észlelési módszerek az automatizált emberi tevékenység felismerésben

DOI 10.17047/Hadtud.2022.32.E.270

A katonai, rendőrségi és bűnüldözési szervezetek hatalmas mennyiségű adattal dolgoznak, de hiányzik az információk elemzéséhez és osztályozásához szükséges humán erőforrás. A biometrikus felismerési alkalmazások, például az arcfelismerés egyre szélesebb körben elterjedt, hogy a hatóságok a biztonság hatékonyságának növelése érdekében alkalmazzák a technológiát. A járvány miatt az arcmaszk viselése számos országban törvényi előírássá vált. Tanulmányok kimutatták, hogy az arcfelismerés pontossága jelentősen csökkent az arcmaszkok miatt. A cél egy új, mesterséges intelligencián és mély neurális hálózat tanításán alapuló kutatási irány, nevezetesen az emberi tevékenység felismerése (HAR) és az ilyen automatizált rendszerek megkerülésére szolgáló hamisítási lehetőségek bemutatása.
Kulcsszavak: mesterséges intelligencia, számítógépes látás, emberi tevékenység felismerése (HAR).

Real-world spoofing detection for human activity recognition applications

Military, Police and law enforcement units have a vast amount of data to work with, but they lack the human resource to analyse and classify information. Biometric recognition applications, such as face recognition becomes more and more widespread for the authorities to apply the technology for increasing security efficiency. However, the pandemics influenced wearable objects covering the face, as wearing a face mask became a requirement by law in many countries. Studies revealed that the accuracy of face recognition significantly dropped because of the face masks. The objective is to present a novel field of research based on artificial intelligence and deep neural network teachings, namely, Human Activity Recognition (HAR) and the spoofing possibilities to bypass such automised systems.
Keywords: artificial intelligence, computer vision, human activity recognition (HAR)

1. Bevezetés

A mindenütt jelenlévő számítógépek és intelligens viselhető érzékelők rendelkezésre állásával az emberi tevékenység felismerése (HAR – Human Activity Recognition) az elmúlt évek egyik népszerű kutatási témájává vált. A HAR-algoritmusok célja, hogy az emberek egyszerű vagy összetett fizikai tevékenységéről információkat nyerjenek. Az ilyen algoritmusok különböző érzékelők adatait használják bemenő jelként és gépi tanulást vagy számítógépes látástechnikákat használnak az emberi tevékenységekre vonatkozó információk kinyerésére. Következésképpen a HAR széles körben alkalmazható számos alkalmazásban, pl. orvosi diagnosztikában, idős emberek nyomon követésében[1], intelligens otthonokban[2], automatizált

[ℓ] Nemzeti Közszolgálati Egyetem, Katonai Műszaki Doktori Iskola, email: huszar.viktor.denes@uni-nke.hu; ORCID: 0000-0001-5402-0208.

[1] Jones, S.E.e. 2017.
[2] Jalal, A.; Uddin, M.Z.; Kim, T. 2012, 58, 863–871.

vezetésben[3], katonai kiképzésnél, bűnözői tevékenységek megfigyelése során[4], valamint a mozgás vezérelt virtuális játékokban[5] is.

A HAR-hoz használt érzékelőket külső és viselhető érzékelőkre lehet osztani[6]. A viselhető érzékelők mérik a szükséges adatokat a tevékenység felismeréséhez, miközben fizikai kapcsolatban vannak a felhasználóval. A viselhető érzékelők ismert példái a gyorsulásmérők, giroszkópok és magnetométerek, amelyekkel az emberi mozgást jelmintákká alakítják az aktivitás felismerése érdekében.[7] A külső érzékelők esetében az érzékelőket fix pontokon helyezik el és a felhasználónak interakcióba kell lépnie velük (ez olyan alkalmazásokban gyakori, mint pl. a közterületen használt biztonsági kamerák). A mély tanulási módszerek, például a konvolúciós neurális hálózatok (CNN – Convolutional Neural Network) és az ismétlődő neurális hálózatok fejlődésével lehetséges a legkorszerűbb eredmények elérése az érzékelőadatok automatikus tanulásával. Egy, a külső érzékelőket használó HAR mély tanulási módszereit alkalmazó használat esetében a valós idejű észlelési információk lehetővé teszik számunkra, hogy felismerjük az ember aktuális tevékenységét.

A nagy felismerési pontosság ellenére a HAR-rendszerek sok esetben nem képesek megkülönböztetni a valódi és hamisított embereket, pl. a mobil képernyőn vagy a számítógép monitorjain lejátszott, ott megjelenített személyek esetében fordul elő tévesztés. A behatolók sokféle eszközt használhatnak hamisítási támadások indítására. Az egyik leggyakoribb a videók digitális képernyőn történő lejátszása. A HAR-rendszerek érzékenyek a hamisító támadásokra, amelyek becsaphatják ezeket a rendszereket, hogy egy hamis felhasználót valódi felhasználónak ismerjenek fel. Az egyik legfontosabb biztonsági kihívás az, hogy a csalás és a hamisítás felderítését a meglévő rendszereknek megfelelően kell megtervezni. Bárki, aki illetéktelen tevékenységet tervez, innovatív ötletekkel és megoldásokkal rendelkezhet a gépi számítógépes látás kijátszására, annak érdekében, hogy megkerülje a legegyszerűbb biztonsági kapukat, és esetleg kicselezze a gépi látást. Ezért nagyon fontos, hogy a jövőbeli fejlesztések hatékonyan kiszűrjék az ilyen esetleges csalásokat.

A tudományos kihívás az, hogy a meglévő kamerák képességei nem biztos, hogy mindig képesek jó minőségű adatokat szolgáltatni az időjárási és látási viszonyok miatt. Kihívást jelent, hogy nehéz nagy mennyiségű adatot bevonni a CNN-alapú algoritmusok tanítására, és az általános adatvédelmi rendelet (General Data Protection Regulation – GDPR) és egyéb előírások miatt az adatforrás korlátozott lehet. Különösen olyan esetekben áll rendelkezésre kevés adat, mint a katonai és védelmi alkalmazhatóság és megfigyelés, ahol az adatok gyakran minősítettek, és az adatokhoz való hozzáférés természeténél fogva korlátozott.

Ezért a mesterséges intelligencia használatával előre meghatározott kockázati kritériumokon alapuló profilalkotási kompetenciák alakíthatók ki a meglévő telepített kamerák képeinek feldolgozásával és elemzésével. A képelemzés alapja a tárgyanalízis, a

[3] Xing, Y.; Lv, C.;Wang, H.; Cao, D.; Velenis, E.;Wang, F. 2019, 68, 5379–5390.
[4] Revathi, A.R.; Kumar, D. 2013, 29, 983–1009.
[5] SQILLER App. The digital football game 2019.
[6] Lara, O.D.; Labrador, M.A. 2013, 15, 1192–1209.
[7] Nweke, H.F.; Teh, Y.W.; Al-garadi, M.A.; Alo, U.R. 2018, 105, 233–261.

mozgáselemzés stb. Az adatok ellenőrzése szükség esetén elvégezhető az illetékes hatóságok, például a rendőrség, a Nemzetbiztonsági Szolgálat támogatásával. A további kísérleti fejlesztések szakmai követelményeinek és módszertanának meghatározása, valamint a megvalósíthatósági szakaszok eredményeinek feldolgozása a tudományos kutatóműhelyek feladata lehet.

Az automatizált HAR-rendszerek biztonságos használatához hamisítás-észlelési (spoofing) technikákra van szükség a támadások felismerése miatt. A HAR-rendszerekben a felhasználók mindig mozognak, és a mozgás sebességétől függően ez elmosódást okozhat a rögzített videón, ami megnehezíti a hamisítás észlelését az ilyen alkalmazásokban. Ezen kívül a műveletfelismerés összetettségétől függően néha szükség van vizuális adatok továbbítására egy szerverre a távoli feldolgozás céljából. Ez magában foglalhatja a videó tömörítését és a videó átméretezését, amelyek rontják a videó minőségét, és még nehezebbé teszik a hamisítás észlelését. Ezeknek a problémáknak a kezelésére jelen írás mély tanuláson alapuló kutatási lehetőségeket mutat be.

A tanulmány tudományos értékei

- Mély tanuláson alapuló eljárás, amellyel felderíthetők az embereket rögzítő videókeretekből származó hamisítási támadások. Az algoritmus nagyon pontos és robusztus minden átméretezési és adatfolyam-feldolgozási adattal szemben, miközben továbbra is elég forrás-hatékony ahhoz, hogy mobileszközön futtassa a fő HAR-algoritmusok mellett.
- Új felismerési stratégia a javasolt mélytanulási hálózat észlelésének kombinálására, időlegesen egy rögzített videón vagy egy élő videófolyamon, a szisztematikus videó lejátszás-hamisítási támadások észlelése érdekében, miközben megőrzi a felismerési alkalmazás a valós idejű teljesítményt.
- A javasolt eljárás teljesítményének további értékelése a biometrikus felismerési alkalmazások összefüggésében.

A tanulmány kiterjeszthető a különféle HAR alkalmazásokkal való munkavégzésre, amelyek sebezhetőek a különböző biztonsági alkalmazásokban lejátszott videó lejátszásokból eredő hamisításokkal szemben. A kutatás jelenlegi fázisában azonban csak azokban az esetekben alkalmazható, amikor az emberi arc teljesen vagy részben látható.

2. Kapcsolódó munkálatok

Rengeteg technika áll rendelkezésre a visszajátszási (replay) támadások vizuális adatokból történő kimutatására. Ezek többsége a hamisítás észlelésére vonatkozik a biometrikus felismerési alkalmazásokban, túlnyomórészt arckép-elemzéssel. Ilyenkor a felhasználók a kamera közelében vannak és a kamera felé néznek. A kutatási rendszer viszont először a HAR összefüggésében vizsgálja a videókból származó hamisítás észlelését, ahol a felhasználók távol vannak a kamerától, és nem mindig nézik a kamerát.

A visszajátszási (replay) támadások észlelése területén a szakirodalomban javasolt technikákat négy nagy csoportba lehet sorolni: felhasználói viselkedésmodellezés, felhasználói együttműködés, hardveralapú és adatvezérelt jellemzés.[8]

A viselkedésmodellezési technikák célja a felhasználói műveletek nyomon követése, például a fej mozgása és a szem pislogása. Arra használják, hogy megértsék, ha a kamera előtt álló személy valódi felhasználó, vagy csak egy fénykép róla.[9] Az ilyen módszerek mozgásnagyítással érzékelik az élő emberi arc finom mozgásait.[10] Ezek nem relevánsak a jelenlegi írás szempontjából, mivel csak akkor hatékonyak, ha a támadó fényképet használ. Ezzel szemben a fő cél a videó visszajátszási (replay) támadások felismerése.

A felhasználói együttműködésen alapuló technikák úgy keresnek prezentációs támadásokat, hogy kizárólag másodlagos interakciót kezdeményeznek a felhasználó és az észlelési modul között.[11] Ilyen technikára jó példa az, amikor az alkalmazás arra ösztönzi a felhasználót, hogy bizonyos mozgást vagy műveletet hajtson végre. Az automatizált videómegfigyeléssel kapcsolatos alkalmazásokban nem célszerű ezeket a felhasználói együttműködésen alapuló technikákat használni, mivel előfordulhat, hogy a felhasználók nem szándékosan lépnek kapcsolatba a rendszerrel.

A hardveralapú technikák további hardvereket, például infravörös vagy mélységérzékelőket használnak a mélységinformációik pontosabb feldolgozásához.[12] Az ilyen technikákon belül a mélységjelzések lehetővé teszik, hogy különbséget tegyünk egy lapos tárgy, például egy mobiltelefon képernyője, vagy egy valódi 3D-s tárgy, például az ember között. Így ezek a módszerek nagyobb robusztusságot biztosítanak a megvilágítás és a felhasználó helyzetének változásaival szemben. Az olyan alkalmazásokban azonban, mint például az okostelefonokkal való együttműködésre tervezett mozgásvezérelt játékok, a felhasználók nem mindig rendelkeznek ilyen kiegészítő hardverrel a készülékükön, így ezen módszerek alkalmazási lehetőségei korlátozottak.

Végül vannak olyan technikák, amelyek adatvezéreltek, és jellemzőek az újra játszási támadások esetén, mivel megjósolják/megtanulják a támadási kísérletek jellemzőit a szabványos adatgyűjtő érzékelőből származó vizuális adatok segítségével. Ez a kutatás az ilyen adatvezérelt technikákra összpontosít, rögzített vagy mobil kamerákból nyert vizuális adatok felhasználásával.

2.1. Mélységi elemzés

Létezik egy módszer a hamisítás észlelésére, amely a kép mélységének elemzésén alapul optikai-flow (optical-flow estimation) becslés segítségével. A módszer célja, hogy megbecsülje a felhasználó arcának 3D szerkezetét, hogy különbséget tegyen a 3D élő arc és a

[8] Bresan, R.; Pinto, A.; Rocha, A.; Beluzo, C.; Carvalho, T. FaceSpoof Buster 2019.
[9] Bharadwaj, S.; Dhamecha, T.I.; Vatsa, M.; Singh, R 2013, 105–110.
[10] Zhang, Z.; Yan, J.; Liu, S.; Lei, Z.; Yi, D.; Li, S.Z. 2012, 26–31.
[11] Bao, W.; Li, H.; Li, N.; Jiang, W. 2009, 233–236.
[12] De Marsico, M.; Nappi, M.; Riccio, D.; Dugelay, J.L. 2012, 73–78.

2D hamisított arc között.[1314] A mélységkamera alapú hamisítás észleléshez hasonlóan az élő arcok 3D objektumok, és egyértelműen megkülönböztethetők a 2D sík közegtől, például a fényképektől. Az ilyen módszerek gyakorlatilag használhatók sík statikus közegek támadásának azonosítására. Ha azonban a kamera nem statikus, vagy ha a felhasználó a képen mozog, a megbízható mélységi információk beszerzése a hamisítás észleléséhez nagyon nagy kihívást jelenthet. Ráadásul a mélység- vagy alakelemzés több képkockát vizsgál egyetlen előrejelzés elkészítéséhez, ami ezeket az eljárásokat lassúvá és erőforrás-igényessé teszi.

2.2. Textúraelemzés

A hamisításhoz használt eszközök, mint például a papír vagy a digitális képernyő, eltérő tükrözési tulajdonságokkal rendelkeznek, mint a valós és az élő arcok. A textúraelemzési módszerek ezt a megfigyelést használják fel hamisítás észlelésére.[15] A megvilágított és a fényvisszaverő felületek közötti kölcsönhatás modellezésével lehetséges az albedo (egy felületről visszaverődő diffúz színtérkép) és a normál színtérképek kinyerése[16], amelyek tulajdonságai használhatók a valódi és a hamisított minták megkülönböztetésére.

Ezek a módszerek rövidebb válaszidővel rendelkeznek, mint a mélységi elemzésen alapuló eljárások, mivel gyakran egy-egy képkockát vizsgálnak. Mindazonáltal gyengén képesek általánosítani a HAR-alkalmazások esetében, mivel a valódi videókban tükröződő tárgyak is lehetnek, amelyek becsaphatják az ilyen eljárásokat. Továbbá a HAR-alkalmazások esetén a személyek szétszéledhetnek a képen, a megvilágítás pedig ellenőrizhetetlen. Tehát az alapvető feltételezések a gyakorlatban nem érvényesek. Más textúra alapú módszerek a ’leírásokat’ használva próbálják rögzíteni a nagyfrekvenciás információkat, például a helyi bináris minták (Local Binary Patterns – LBP) változatai.[17]

2.3. Képminőség

A képminőségen alapuló módszerek megbecsülik a fényképek vagy képernyők újbóli rögzítése során bekövetkező általános képminőségromlást. A képminőség romlásához hozzájáruló tényezők közé tartozik az elmosódottság és a színdeformáció.[18192021] A képminőség kiszámításához gyakran szükség van egy referenciaképre, amely nem áll rendelkezésre. A hagyományos eljárás az, hogy a forráskép egy degradált változatát szimuláljuk, és azt a forrásképpel együtt használjuk a képminőség kiszámításához. Itt a hipotézis az, hogy a forrás- és a szimulált kép közötti objektív minőségromlás a valós képek esetében viszonylag kisebbek, mint a támadó képek esetében. Így, ha a képfelvételi feltételek

[13] Bao, W.; Li, H.; Li, N.; Jiang, W. 2009, 233–236.
[14] De Marsico, M.; Nappi, M.; Riccio, D.; Dugelay, J.L. 2012, 73–78.
[15] Chingovska, I.; Anjos, A.; Marcel, S. 2012, 1–7.
[16] Tan, X.; Li, Y.; Liu, J.; Jiang, L. 2010, 504–517.
[17] Boulkenafet, Z.; Komulainen, J.; Hadid 2016, 11, 1818–1830.
[18] Galbally, J.; Marcel, S.; Fierrez, J. 2014: 23, 710–724.
[19] Unnikrishnan, S.; Eshack 2016, 1–5.
[20] Garcia, D.C.; de Queiroz, R.L. 2015, 10, 778–786.
[21] Arashloo, S.R.; Kittler, J. 2017, 5, 13868–13882.

nem hasonlóak, ez a hipotézis nem érvényesül, ami ezeket a módszereket érzékennyé teszi a HAR-forgatókönyvekkel szemben.

2.4. Frekvenciatartomány elemzés

A digitális hamisító adathordozókon történő felvételek a bemutató adathordozók diszkréciója miatt nagyfrekvenciás zajt visznek be a képadatokba. Ezt a gyakoriságra vonatkozó információt Fourier-elemzéssel rögzíthetjük.[22] A frekvenciatartományon alapuló módszerek feltárják ezeket a zajjeleket a rögzített videókban, hogy megkülönböztessék az élő és a hamis arcokat.[23] Ezek a zajjelek a frekvenciatartományban erős jelek a támadások észlelésére. Meg kell azonban jegyezni, hogy a nagy felbontású hamisító médiumok használata csillapíthatja ezt a zajt, és a felismerhető zajminták nem mindig vannak jelen, ami miatt az ilyen eljárások kizárólag azokra alapozva nem megbízhatóak.

2.5. Mély tanuláson alapuló módszerek

Az utóbbi időben számos képfelismerő alkalmazással összefüggésben bebizonyosodott, hogy a CNN-re épülő modellek a leghatékonyabb teljesítményt érték el. Az ilyen mély tanulási eljárások megtanulják a köztes ábrázolások előrejelzését közvetlenül a pixeladatokból anélkül, hogy a teljes adatmintákból kinyerhető, komoly számítási kapacitásokat felhasználó eredményekre hagyatkozzanak. A szakirodalom nem taglalja a CNN-ek használatát a hamisítás észleléshez HAR-alkalmazásokban, azonban más architektúrákat javasolnak a hamisítás észlelésére, különösen az arc biometrikus hitelesítési alkalmazásaiban lévő videóknál. Rodrigo B. és munkatársai[24] javasoltak egy módszert a ResNet50 használatával.[25] Ezt a neurális hálózatot több előre kiszámított érték, például mélység-, rugalmasság- és megvilágítási értékek segítségével tanították, ami a módszertani kontextustól is függ. Az ilyen eljárások használata mobileszközön nehézkes lehet, mivel ezeknek az értékeknek a kiszámítása rendkívül erőforrás-igényes lehet.

Yaojie L. et. al.[26] a CNN-ek és a visszatérő neurális hálózatok (Recurrent Neural Network- RNN-ek) kombinációját alkalmazó módszert javasolt a hamis arcok észlelésére. Különösen a képek RGB (Red, Green, Blue) + HSV (Hue, Saturation, Value) reprezentációját használják a kutatásukban, és több videókeretet használnak egyetlen előrejelzéshez, azonban ezek a hálózatok általában lassabbak, mivel több képkockán működnek.

Atoum. Y. et. al.[27] bevezette a két streames CNN-t, amely kiszámítja a helyi jellemzőket a javításokból és a mélységi értékeket, majd egyesíti ezeknek a folyamoknak a kimenetét a visszajátszási (replay) támadások észlelése érdekében. A helyi szolgáltatások használata robusztussá teszi ezeket a módszereket. A mélységszámítás miatt azonban előfordulhat, hogy

[22] Li, J.; Wang, Y.; Tan, T.; Jain, A.K. 2004, 5404, 296–303.
[23] Pinto, A.; Pedrini, H.; Schwartz, W.R. 2015: 24, 4726–4740.
[24] Bresan, R.; Pinto, A.; Rocha, A.; Beluzo, C.; Carvalho, T. FaceSpoof Buster 2019.
[25] He, K.; Zhang, X.; Ren, S.; Sun, J. 2015.
[26] Liu, Y.; Jourabloo, A.; Liu, X. 2018.
[27] Atoum, Y.; Liu, Y.; Jourabloo, A.; Liu, X. 2017, 319–328.

nem tudnak lépést tartani a valós idejű követelményekkel, különösen akkor, ha egy tevékenységfelismerő algoritmusnak párhuzamosan kell futnia mobileszközön is. A jelenlegi írásban a mobileszköz alkalmazás kerül bemutatásra, ahol a csekély forrásigényű modellek nagy előnyt élveznek.

A hamisítás észlelésének fent említett technikái jól működnek a mesterséges környezetekben, a világítás és a háttér speciális beállításával. Szükséges továbbá, hogy a felhasználóknak szembe kell nézniük a kamerával az észlelés végrehajtásához, amely intuitív a biometrikus felismerési alkalmazások számára. A gyakorlatban a fenti megoldások nem terjedhetnek ki a HAR-alkalmazásokra, mivel a felhasználók rugalmasan választhatják a helyüket, és végül nem mindig lehet ellenőrizni a felhasználói viselkedést. Ilyen kiszámíthatatlan emberi tevékenység és mozgásminta előfordulhat az egyetemeken, más intézményekben és katonai helyszíneken. Ezenkívül bizonyos esetekben, például virtuális játékokban egyes testrészeket, beleértve az arcot is, részben vagy teljesen eltakarják az interakciós tárgyak. A hagyományos hamisítás észlelési módszerek a biometrikus arc-hitelesítési alkalmazások esetében nem terjedhetnek ki az ilyen esetekre. Ráadásul, ha a tevékenységfelismerés távoli szerverfeldolgozást foglal magában, a képadatok átméretezhetők és/vagy tömöríthetők. Az ilyen műveletek függését a hamisítás észlelési pontosságától a szakirodalom nem tárgyalja. Mindezen problémák megoldása érdekében V. Kirannal közösen a Sensors-ban publikált kutatásban új eszközöket dolgoztunk ki, beleértve egy új adatbázist és egy adatközpontú eljárást, amelyek közösen megvizsgálják a rögzített kép különböző régióit a hamisítás észlelése érdekében. A rendszer fejleszti és integrálja a legkorszerűbb mélytanulási algoritmusokat a HAR-alkalmazások replay támadási hamis eseteinek észlelésére. [28]

3. Vizuális elemzés a videó lejátszás hamisításának észleléséhez

Egy CNN-modellen alapuló rendszer, amely megkülönbözteti a valódi személyeket a hamisított személyektől, vizuálisan elemezhető, azonban magasabb biztonsági igényű alkalmazási helyeken, például katonai védett területeken vagy a határvédelemben és a határellenőrzésben szükséges már a publikációban ismertetett eljárás használata, hiszen pontosabb csalás felismerést igényelnek.

3.1. Kontextus kiválasztása a hamisítás észleléséhez

Általában a felismerendő felhasználói tevékenységtől és annak összetettségétől függően a tevékenységfelismerő algoritmusok egy vagy több testrész/ízület helyének kiszámítását, vagy bizonyos esetekben a felhasználó teljes csontvázának képből történő követését foglalják magukban.[29] Az interakciós tárgy (pl. egy tárgyat használó virtuális játék) esetében a felhasználó testrészeivel együtt a felhasználó helyét is ki kell számítani a képen, mivel itt a tevékenység a felhasználó és a tárgy interakcióját jelenti.

[28] Huszár, V.D.; Adhikarla, V.K. 2021, 21, 7339.
[29] Cao, Z.; Hidalgo, G.; Simon, T.;Wei, S.E.; Sheikh, Y. OpenPose. 2019.

A kutatásokhoz szükséges egy adatbázis (data set) létrehozása, és a címkék összegyűjtése. Egy ilyen ’alkotás’ után betanításra kerül a YOLO[30] mély tanulási CNN architektúra, és később ez kerül alkalmazásra az emberi póz és egy adott objektum, tárgy együtt történő felismerésére a videóképekből egyetlen felvételen. A YOLO gyors és robusztus teljesítményt nyújt még mobileszközökön is.

Mivel a biometrikus hitelesítési rendszereket széles körben használják a védelmi szférában, számos biztonsági alkalmazásban, ezért a hamisítás észlelésével folytatott további kutatás rendkívül fontos lesz. A kutatásoknak figyelniük kell azon esetekre is majd, amikor az emberi test teljes egészében nem látható.

3.2. Modellek

A videókkal való munka közben többféleképpen kombinálhatóak az időbeli információk, ami megnehezíti a rögzített méretű architektúra használatát. A kutatás során a videók több rövid klipként kerülnek kezelésre, amelyek azonos méretű képkockákból állnak. Az ötlet az, hogy a klip ezen több képkockáját felhasználva megtanuljuk a térbeli-időbeli jellemzőket.
Egy ilyen rendszert érdemes a Keras[31] programban implementálni. Ebben kerülhetnek vizsgálhatóvá a nyomvonalak tér-időbeli jellemzőinek tanulásai.

3.2.1. Egykeretes modell (Single frame model - SF)

A statikus keretekben rejlő lehetőségek tanulmányozásához a valódi és hamis esetek pontos osztályozásában ez a modell kerülhet a kutatás középpontjába. Az SF felveszi a videó minden képkockáját, és megjeleníti a megfigyeléseket. Itt a jelenlegi bemeneti Full HD képkockából kibontott archatároló doboz átméreteződik 64X64X3 képpontra, majd betáplálásra kerül egy CNN-hez.

3.2.2. Összekapcsolt keretmodell (Concatenated frames model - CF)

A módszer mögötti elgondolás alapja az, hogy a vizuális információk egy videoklipben kerülnek egyesítésre az összefüggő képkockákat tartalmazó időablakban. Ez úgy kerül elérésre, hogy a szűrők az alapmodell első VGG blokkjának konvolúciós rétegén kerülnek adaptálásra. Az SF-hez hasonlóan az egymást követő képkockák a figyelembe vett időablakban 64X64X3 képpontra kerülnek átméretezésre, és a kezdeti konvolúciós szűrők 64X64X3XN képpont méretűre kerülnek bővítésre, ahol N a figyelembe vett időablak mérete. A Sensorsban ismertetett vizsgálat során N = 5 került kiválasztásra a helyi minták megismerésére és észlelésére az időablakban.

[30] Redmon, J.; Farhadi 2018.
[31] Chollet, F.; others 2015.

3.2.3. Késleltetett keretek modellje (Delayed frames model - DF)

A késleltetett képkockák modellje két különálló alapmodellt használ a második VGG blokkig, és ezt a két modellt két 64X64X3 átméretezett képkocka táplálja, amelyek P képkockányira vannak egymástól az adott videósorozatban. A második VGG blokk után e két modell kimeneteit egyesítik, és az alapmodell többi, teljesen összekapcsolt rétegéhez csatlakoztatják. A kutatás során a P = 15 érték kerülhet használatra a globális jellemzők megismerésére és észlelésére.

3.2.4. Ensemble multi-stream model (EM)

Az Ensemble modell az átméretezett fejhatároló dobozt vizsgálja három különböző feldolgozási folyamban, három térbeli felbontásban. Az SF modellhez hasonlóan egy-egy keret kerülhet figyelembevételre. A bemeneti HD videóból egyetlen képkocka kerül kibontásra, és a YOLO modellhez kerül, hogy kinyerje a fej fejrészét. Az észlelt fejhatároló doboz átméreteződik 64X64X3 képpontra, és három különböző adatfolyamban kerül feldolgozásra.

Az első folyamban az átméretezett fejkép az alapmodellbe kerülhet betáplálásra (ugyanaz, mint az SF). A második folyamban az átméretezett fejkép alsó 64X32X3 pixelei levágásra kerülnek, és további feldolgozásra, már a módosított alapmodellbe kerülnek betáplálásra. Az alapmodell kezdeti konvolúciós rétegén a szűrők úgy kerülnek módosításra, hogy ebben az adatfolyamban a levágott kép méretének megfelelőek legyenek. Ez az adatfolyam főként az áll környéki vizuális adatokat kapja, és így a szemüveget vagy sapkát viselő felhasználók által okozott torzítás minimálisra csökken. Az utolsó adatfolyamba az átméretezett kép 32X32X3 pixeles, középső 32X32X3 pixeles adatai kerülhetnek. Ezeket az adatokat a legjobb, ha egy másik módosított alapmodell dolgozza fel, amely a szükséges kezdeti konvolúciós szűrők méretével rendelkezik. Az adatfolyam az arc központi kivágott részét dolgozza fel, és így nem rendelkezik a háttérből származó zavaró mintázatok ’zajával’. A három adatfolyamból származó észleléseket a felismerési valószínűségek és a többségi szavazás alapján összevonhatjuk, hogy egy együttes észlelési értéket kapjon a detekció.

4. Eredmények és vita

4.1. Tesztelési séma

Mivel a hamisított esetek folyamatos észlelése a cél a biztonsági rendszerekben, akár rögzített, akár élő videó adatfolyamokon, ezért egymást átfedő videóklipek darabjai kerülnek meghatározásra, amelyek mindegyike több képkockát tartalmaz (a tesztelhető modelltől függően), 15 képkocka átfedéssel. Minden kliphez egyetlen előrejelzés kerül hozzáadásra. Ebből a célból egy klipen belül a modellek három képkockán kerülnek futtatásra, amelyek egymástól 15 képkockát tesznek ki. Ezek az előrejelzések egyesítésre kerülnek, hogy egyetlen előrejelzés kerüljön kapcsolásra ehhez a kliphez. Ezért adott videófolyam esetén az SF és EM modellek esetén az első előrejelzés 31 képkocka után történik, ezt követően pedig 15

képkockán belül történik. A DF modell esetén, mivel bemenetként két képkockát vesz fel, amelyek 15 képkocka távolságra vannak egymástól, az első előrejelzés csak 45 képkocka után történik, ezt követően pedig 15 képkockánként. A CF modell esetén, mivel 5 egymást követő képkockát vesz igénybe bemenetként, az első előrejelzés 35 képkocka után történik, ezt követően pedig 15 képkockán belül.

4.2. Kísérletek az adatmintákon (data set)

Az eddigi ismeretek szerint egy ilyen rendszer a maga nemében először a videókból származó hamisítás észlelését vizsgálja a HAR összefüggésében. Az várható, hogy a vizsgálandó személyek mindig különböző távolságra lesznek a kamerától, s az nem biztos, hogy minden testrészük mindig teljesen látható. A személyek arc orientációja a videók között nagymértékben változik, és néha az arcot teljesen vagy részben eltakarja egy tárgy. Ezek a képek kéz és fej határoló dobozokkal fűződnek a játékosok arc- és háttérinformációihoz, amelyeket egy YOLO modell képzésére használunk, amely valós időben észleli a fejrészeket. A legmagasabb minták száma körülbelül 44 képpont. Az adatbázis olyan mintákat is tartalmaz, ahol az IPD (InterPupillary Distances) nyilvánvalóan 0-50 képpont között helyezkedik el, ami azt mutatja, hogy a felhasználók milyen távolságra vannak a kamerától. A 0 pixeles IPD azt jelentheti, hogy az arc nem teljesen látható, vagy a játékos merőleges irányba néz a kamerával szemben.

4.2.1 Videóklip szintű előrejelzések

A Sensorsban megjelent publikációban ismertetett eredmények a biometrikus prezentációs támadások értékelésében általánosan figyelembe vett mérőszámok – Attack Presentation Classification Error Rate (APCER), Bona fide Presentation Classification Error Rate (BPCER) és Half-Total Error Rate (HTER)[32] – segítségével kerültek közlésre. Az APCER és a BPCER a hamis elfogadási arány (FAR) és a hamis elutasítási arány (FRR) analógjai. Így az APCER, BPCER és következésképpen a HTER alacsonyabb értékei a modell jobb teljesítményét jelzik. Az eredmények közléséhez a modellek kimenetei a kapott értékek szerint kerülnek felhasználásra, és az osztály-előrejelzési valószínűségeken alapuló küszöbértékek alapján nem kerültek alkalmazásra.

A fő megállapítás az, hogy az Ensemble modell jelentős előnnyel felülmúlta a többi módszert aggregált eredmények alapján. A Sensorsban bemutatott kutatási adatminta tesztkészleten a hamis pozitív arány (FPR) és az igaz pozitív arány (TPR) közötti kompromisszum elemzésével is igazolhatja egy kutató, de természetesen más adatminta esetén is ez az eredmény várható. Megjegyzendő, hogy a Single frame módszer is megbízható eredményeket ad. A kísérletek azonban azt mutatják, hogy a képkockák időbeli kombinálása nem hozott jobb eredményeket. Különösen a CF-modell esetében az APCER-érték magas, ami azt jelzi, hogy ez a modell több csaló felhasználót érvényes felhasználóként fogad el. Ez arra utal, hogy a kombinált képkockák nem mindig tartalmaznak olyan mintákat, amelyek a

[32] ISO. ISO/IEC JTC 1 /SC 37 2017

mélytanulási modell által megfelelően jellemezhetőek, és teljesen önkényesek lehetnek. A DF-modell esetében a BPCER-érték magas, ami azt jelzi, hogy számos valódi felhasználót ez a modell hamis felhasználóként kategorizált. Ebből a viselkedésből az a feltételezés következik, hogy a DF modell megtanult jellemzőket, amelyek inkább a mozgásra, mint a hamisítás felismerésével kapcsolatos mintákra vonatkoznak.

4.2.3. Kísérletek videó tömörítéssel

Amint az korábban bemutatásra került, a felismerendő művelet összetettségétől függően a videót egy távoli szerverre kell eljuttatni (pl. streamelni), ahol a videófolyam tényleges feldolgozása történik. Az olyan esetekben, mint a mozgás vezérelt virtuális játékok, ahol a felhasználói videókat egy mobiltelefonról egy távoli szerverre továbbítják, nem mindig van elegendő hálózati sávszélesség a videó natív minőségben vagy magas felbontásban történő adattovábbításához. Ezt követően videó tömörítési technikákat alkalmaznak a videó bitrate / sebesség csökkentésére. Annak értékelése érdekében, hogy az EM modell képes-e helyesen osztályozni a hamis eseteket, tömörített videófolyamokat állítottunk elő, amelyek bitsebessége 300 kbps és 1500 kbps között változik.

Egy ilyen kísérlethez egy rögzített videó esetén az értékelést teljesen offline lehetséges elvégezni. Az eredmények azt mutatják, hogy az együttes modell robusztusan teljesít extrém tömörítés (300 kbps) esetén is.

4.3. Kísérletek az arcfelismerő adatbázisokkal

Annak értékeléséhez, hogy a hamisítás-felismerő modell képes-e általánosítani más területekre, például arcfelismerő rendszerekre, két széles körben ismert adatkészlet kerülhet bevonásra a kutatásba: az Idiap REPLAY-MOBILE[33] és CASIA Face AntiSpoofing.[34] A CASIA Face AntiSpoofing adatbázis 50 alany 600 videoklipjéből áll. A 600 videoklipből 150 videoklip videó-visszajátszási (replay) támadást jelent. Az Idiap adatbázissal összehasonlítva a CASIA DB különböző kamerákkal (Sony NEX-5-HD, két alacsony minőségű egyéb USB-s kamera) rögzített képeket tartalmaz, amelyek egy iPad-en megjelenített visszajátszási támadásokat rögzítenek. Ennek az adatbázisnak azonban jelentős hiányossága, hogy a videós visszajátszási támadásokat nagyon alacsony felbontásban rögzítették (640x480). A REPLAY-MOBILE adatállomány 1190 videoklipből áll, amelyek 40 ügyfélnek különböző megvilágítási körülmények között végrehajtott fénykép- és videóbemutató támadások (spoofing attacks) videoklipjeit tartalmazzák. Ezeket a videókat egy (iOS-t futtató) iPad Mini2 és egy LG-G4 okostelefonnal rögzítették, teljes HD felbontásban.

Az 1. táblázat a figyelembe vett arcfelismerő adatbázisok tesztkészleteinek értékelési eredményeit mutatja. Az arcképek az adatbázisból nyert kiegészítő adatok felhasználásával kerültek levágásra, majd 64x64 képpontra méretezésre, mielőtt betáplálásra kerültek az EM

[33] Costa-Pazo, A.; Bhattacharjee, S.; Vazquez-Fernandez, E.; Marcel, S. 2016, 1–7.
[34] Zhang, Z.; Yan, J.; Liu, S.; Lei, Z.; Yi, D.; Li, S.Z. 2012, 26–31.

modellbe. Az eredmények azt mutatják, hogy az eljárás jól teljesít a REPLAY-MOBILE adatbázisban, függetlenül attól, hogy a felhasználók nagyon közel vannak a kamerához (magasabb IPD értékek, mint az adatbázisunké). A Sensorsban bemutatott módszer viszonylag jól teljesít a CASIA adatbázisban is. A relatív teljesítmény csökkenés a CASIA adatbázis esetében az alacsony felbontású bemeneti képeknek köszönhető. Fontos megjegyezni, hogy a módszer a Full HD bemeneti képekből kinyert képadatok alapján kerültek betanításra azzal a feltételezéssel, hogy a tanult funkciókészletünk nem nagyon felel meg az alacsonyabb felbontású képeknek. Ez nem jelent egyébként problémát, mivel a HAR-alkalmazásokban használt jelenlegi generációs érzékelők, beleértve a felügyeleti rendszereket, képesek teljes HD képeket rögzíteni.

Módszer	APCER	BPCER	HTER
SF	9.9010	12.9496	11.4253
CF	36.4486	1.7007	19.0746
DF	3.1915	45.5556	24.3735
EM	8.9109	6.1151	**7.5130**

1. táblázat

Tesztelési adatbázis eredménye

Forrás: Huszár, V.D.; Adhikarla, V.K. „Live Spoofing Detection for Automatic Human Activity Recognition Applications.” Sensors (2021): 21, 7339. https://doi.org/10.3390/s21217339.

5. Következtetések és jövőbeli munka

A kutatás során a videó visszajátszás hamisítás felismerése került tanulmányozásra olyan alkalmazásokban, ahol az emberi tevékenység felismerése RGB-érzékelők segítségével döntő szerepet játszik. A kültéri időjárási és látási viszonyok nem ellenőrizhetők, ezért a hamisítás felismerése ilyen körülmények között bonyolult feladat. Megfogalmazásra került egy olyan ensemble multi-stream modell, amely felismeri a videó visszajátszási támadásokból eredő hamisítási eseteket. A modell az emberi arc különböző régióit vizsgálja, és ezeket a megfigyeléseket kombinálja, hogy robusztus osztályozást biztosítson a hamisított és a valódi esetek között.

Egy ilyen modell akkor is megbízható eredményeket ad, ha az alany arca csak részben látható. Értékelésre került továbbá a betanított modell teljesítménye különböző bitrátájú tömörített videókon, valamint az arcfelismerő rendszerekre való általánosítás képessége is. Az arcfelismeréshez azonban szükség lehet a modell finomhangolására, hogy jobban illeszkedjen az olyan esetekhez, ahol az IPD teljes HD felbontásnál magasabb, mint az eredeti képeknél. Az algoritmus mobileszközön a megvalósításra és a hitelesítésre is alkalmazható, valamint bemutatásra került, hogy az eljárás valós időben, minimális memóriaigény mellett működhet, így elegendő erőforrás kapacitás marad a tevékenységfelismerő algoritmus melletti futtatásra. A jövőben a módszer könnyen adaptálható olyan mobileszközökre, amelyek fejlett érzékelőket tartalmaznak, mint például a mélységérzékelők, és ez tovább javítaná a hamisítás-felismerő eljárásunk teljesítményét. A jövőben a visszajátszási támadások mellett az arcot

és/vagy más testrészeket takaró maszkok vagy kapucnis sapkák és/vagy egyéb ruhadarabok felismerésére szolgáló algoritmusokat is tanulmányozni és feltárni szükséges, amelyeket a hamisítás-felismerő rendszerek megzavarására terveztek.

FELHASZNÁLT IRODALOM

Atoum, Y.; Liu, Y.; Jourabloo, A.; Liu, X. 2017. Face anti-spoofing using patch and depth-based CNNs. 319–328. *IEEE International Joint Conference on Biometrics (IJCB)*.

Arashloo, S.R.; Kittler, J.; 2017. Christmas,W. An Anomaly Detection Approach to Face Spoofing Detection: A New Formulation and Evaluation Protocol. *IEEE Access* 5

Bao, W.; Li, H.; Li, N.; Jiang, W. 2009. A liveness detection method for face recognition based on optical flow field. 233–236 *International Conference on Image Analysis and Signal Processing.*

Bharadwaj, S.; Dhamecha, T.I.; Vatsa, M.; Singh, R. 2013. Computationally Efficient Face Spoofing Detection with Motion Magnification. 105–110. *IEEE Conference on Computer Vision and Pattern Recognition Workshops.*

Boulkenafet, Z.; Komulainen, J.; Hadid, A. 2016. Face Spoofing Detection Using Colour Texture Analysis. 11, 1818–1830. IEEE Transactions on Information Forensics and Security. https://doi.org/10.1109/TIFS.2016.2555286.

Bresan, R.; Pinto, A.; Rocha, A.; Beluzo, C.; Carvalho, T. FaceSpoof Buster 2019. Presentation Attack Detector Based on Intrinsic Image Properties and Deep Learning.

Cao, Z.; Hidalgo, G.; Simon, T.;Wei, S.E.; Sheikh, Y. OpenPose. 2019. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. [arXiv:cs.CV/1812.08008].

Chingovska, I.; Anjos, A.; Marcel, S. 2012. On the effectiveness of local binary patterns in face anti-spoofing. 1–7. *BIOSIG – Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG).*

Chollet, F.; others2015. „Keras.” https://github.com/fchollet/keras.

Costa-Pazo, A.; Bhattacharjee, S.; Vazquez-Fernandez, E.; Marcel, S. 2016. The Replay-Mobile Face Presentation-Attack Database. 1-7. *International Conference of the Biometrics Special Interest Group (BIOSIG).*

De Marsico, M.; Nappi, M.; Riccio, D.; Dugelay, J.L. 2012. Moving face spoofing detection via 3D projective invariants. 73–78. *IAPR International Conference on Biometrics (ICB)*. https://doi.org/10.1109/ICB.2012.6199761.

Galbally, J.; Marcel, S.; Fierrez, J. 2014. Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition. 23, 710–724. *IEEE Transactions on Image Processing*. https://doi.org/doi:10.1109/TIP.2013.2292332.

Garcia, D.C.; de Queiroz, R.L. 2015. Face-Spoofing 2D-Detection Based on Moiré-Pattern Analysis. 10, 778–786. *IEEE Transactions on Information Forensics and Security*. https://doi.org/doi:10.1109/TIFS.2015.2411394.

Jalal, A.; Uddin, M.Z.; Kim, T.. 2012. Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home. 58, 863–871. *IEEE Transactions on Consumer Electronic*s. https://doi.org/10.1109/TCE.2012.6311329.

Huszár, V.D.; Adhikarla, V.K. 2021. *Live Spoofing Detection for Automatic Human Activity Recognition Applications*." Sensors. 21, 7339. https://doi.org/10.3390/s21217339.

ISO. ISO/IEC JTC 1 /SC 37. 2017. Biometrics Information technology – Biometric presentation attack detection – Part 3: Testing and reporting; International Organization for Standardization.

Jones, S.E.e. 2017. Wearable smart sensor systems integrated on soft contact lenses for wireless ocular diagnostics. *Nat. Commun*.. https://doi.org/10.1038/ncomms14997.

Liu, Y.; Jourabloo, A.; Liu, X. 2018. Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision. [arXiv:cs.CV/1803.11097].

Li, J.; Wang, Y.; Tan, T.; Jain, A.K. 2004. Live face detection based on the analysis of Fourier spectra. Biometric Technology for Human Identification. 5404, 296 – 303. *Jain, A.K.; Ratha, N.K., Eds. International Society for Optics and Photonics, SPIE*. https://doi.org/doi:10.1117/12.541955.

Määttä, J.; Hadid, A.; Pietikäinen, M. 2011. Face spoofing detection from single images using micro-texture analysis. 1–7. International Joint Conference on Biometrics (IJCB). doi:10. https://doi.org/1109/IJCB.2011.6117510.

Nweke, H.F.; Teh, Y.W.; Al-garadi, M.A.; Alo, U.R. 2018. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. 105, 233–261. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2018.03.056.

Pinto, A.; Pedrini, H.; Schwartz, W.R.; Rocha, A. 2015. Face Spoofing Detection Through Visual Codebooks of Spectral Temporal Cubes. 24, 4726–4740. *IEEE Transactions on Image Processing*. https://doi.org/doi:10.1109/TIP.2015.2466088.

Redmon, J.; Farhadi, A. YOLOv3: 2018. An Incremental Improvement." *CoRR*. abs/1804.02767, [1804.02767].

Revathi, A.R.; Kumar, D. 2013.A Survey Of Activity Recognition And Understanding The Behavior In Video Surveliance. 29, 983–1009. *Vis Comput*.

SQILLER App. The digital football game 2019. *URL: https://sqillerapp.com/*, Utolsó letöltés: 2021.07.19.

Tan, X.; Li, Y.; Liu, J.; Jiang, L. 2010. Face Liveness Detection from a Single Image with Sparse Low Rank Bilinear Discriminative Model. 504–517. *Computer Vision – ECCV 2010; Daniilidis, K.; Maragos, P.; Paragios, N., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg*

Unnikrishnan, S.; Eshack, A. 2016. Face spoof detection using image distortion analysis and image quality assessment. 1-5. *International Conference on Emerging Technological Trends (ICETT)* https://doi.org/doi:10.1109/ICETT.2016.7873742.

Xing, Y.; Lv, C.;Wang, H.; Cao, D.; Velenis, E.;Wang, F. 2019. Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach. 68, 5379–5390. *IEEE Transactions on Vehicular Technology*. https://doi.org/10.1109/TVT.2019.2908425.

Zhang, Z.; Yan, J.; Liu, S.; Lei, Z.; Yi, D.; Li, S.Z. 2012. A face antispoofing database with diverse attacks. 26–31. *IAPR International Conference on Biometrics (ICB)*. https://doi.org/doi:10.1109/ICB.2012.6199754.

Zhang, Z.; Yi, D.; Lei, Z.; Li, S.Z. 2011. Face liveness detection by learning multispectral reflectance distributions. 436–441. *IEEE International Conference on Automatic Face Gesture Recognition (FG)* https://doi.org/doi:10.1109/FG.2011.5771438.