

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
GÉPÉSZMÉRNÖKI KAR
MECHATRONIKA OPTIKA ÉS GÉPÉSZETI INFORMATIKA TANSZÉK



ORSZÁGOS TUDOMÁNYOS DIÁKKÖRI KONFERENCIA 2023

FABRINCZIUS ZORÁN GYÖRGY, GÖRCES ANDRÁS,

HUNYADY BOTOND, SZABÓ ANDRÁS ZORÁN

(MECHATRONIKAI MÉRNÖK BSC, 2021. ÉVFOLYAM)

Emberi és számítógépes mozgásfelismerés összefüggése- inek vizsgálata

Konzulensek:

Alen Hajnal, PhD

Full Professor (University of Southern Mississippi)

Dr. Samu Krisztián

egyetemi docens (BME MOGI)

Budapest, 2022.

NYILATKOZATOK

Elfogadási nyilatkozat

Ezen dolgozat a Budapesti Műszaki és Gazdaságtudományi Egyetem Gépészmérnöki Kara által a Tudományos Diákköri Konferenciára előírt valamennyi tartalmi és formai követelménynek, továbbá a célkitűzésekben megfogalmazottaknak maradéktalanul eleget tesz. Ezen dolgozatot a nyilvános bírálatra és nyilvános előadásra alkalmasnak tartjuk.

A beadás időpontja: 2022.11.07.

Alen Hajnal és Samu Krisztián
konzulensek

Nyilatkozat az önálló munkáról

Alulírott, Fabrinczius Zorán György (NQTP73), Görcs András (PKDICK), Hunyady Botond (M2HGE1), Szabó András Zorán (UT9O7A), a Budapesti Műszaki és Gazdaságtudományi Egyetem hallgatói, büntetőjogi és fegyelmi felelősségem tudatában kijelentjük és sajátkezű aláírással igazoljuk, hogy ezt a dolgozatot meg nem engedett segítség nélkül, saját magunk készítettük, és dolgozatunkban csak a megadott forrásokat használtuk fel. Minden olyan részt, melyet szó szerint vagy azonos értelemben, de átfogalmazva más forrásból átvettünk, egyértelműen, a hatályos előírásoknak megfelelően, a forrás megadásával megjelöltük.

Budapest, 2022.11.07.

**Fabrinczius Zorán György, Görcs András,
Hunyady Botond és Szabó András Zorán**
hallgatók

ÖSSZEFOGLALÓ

Emberi és számítógépes mozgásfelismerés összefüggéseinek vizsgálata

Az emberi érzékelés egyik alappillére a látás. Vizuális információ segítségével vagyunk képesek saját, de főleg környezetünk mozgásának felismerésére. Pusztán vizuális információból a mozgások felismerése nem triviális. Egy nemrég elvégzett humán kutatásban (Masoner, 2022) a kísérleti alanyoknak egy testre rögzített kamera által készített videófelvevételeket kellett megtekinteni. A kutatók bebizonyították, hogy pusztán a kameramozgás (a mozgó test látványa nélkül) is elegendő információt szolgáltatott a testmozgás felismeréséhez. A kutatásból nem derült ki, hogy milyen fajta vizuális mintázat szolgáltatja az információt. Gibson (1979) hipotézise szerint a mozgásfelismerés alapja az optic flow, ami a vizuális ingert létrehozó fényviszonyok időbeli változását térképezi fel a látható tér minden egyes pontjában.

A jelenlegi kutatás célja az említett humán kísérleti munka eredményeinek (és Gibson hipotézisének) igazolása az ott felhasznált videók számítógépes elemzésével. A videókban a képpontok intenzitásának a változását számítottuk ki képkockák között, ami gyakorlatilag a kamera környezethez viszonyított sebességét mutatja. A kapott sebesség-idő függvényekkel definiáltuk a mozgásokat, illetve vizsgáltuk a videóban szereplő mozdulatok hasonlóságát, egymással való korrelációját. Az összehasonlítás alapjául az intenzitás függvények korrelációját vettük, illetve az ún. dynamic time warping (DTW) paramétert számoltuk ki.

A kapott eredmények alátámasztják a korábbi kutatás eredményeit: a függvények korrelációjának mértéke összefügg a humán kísérletben résztvevők által elért eredménnyel: minél jobban hasonlít két mozgás függvénye annál kevesebben tudták megkülönböztetni a mozgásról készült egocentrikus videókat. A videóelemzés eredményei egyrészt hozzájárulnak a humán észlelés alapkutatásához a

mozgásfelismerés alapjául szolgáló vizuális információ feltérképezése által, másrészt pedig a gépi látás rendszerek precizitását is növelik, ami alkalmazható a rendőrök által viselt testre rögzített kamera felvételek automatikus elemzéséhez.

ABSTRACT

Assessing connections between human and computer-based motion recognition

Vision is one of the most important aspects of human perception. We are able to perceive the movement of our own body and of the environment based on visual information. Recognizing and telling apart motion based solely on visual information is far from trivial. In a recent human perception study (Masoner, 2022), participants were asked to watch videos recorded by a body-mounted camera. The researchers had proven that the video in itself, without direct view of the body, provides sufficient information for identifying motion. This research, however, did not reveal what kind of optical pattern carries this visual information. As hypothesized by Gibson (1979), motion recognition is based on optic flow, defined by the temporal changes of light patterns in the ambient optic array that serves as the stimulus for perception. The goal of the current study was to review and corroborate the results of the visual perception experiments (and Gibson's hypothesis) through computer analysis of the videos previously used.

In our study we calculated the change in pixel intensity between frames, which indicates the velocity of the camera relative to the environment. The obtained velocity-time functions were used to define movements. These functions were compared and assessed for similarities. The numerical comparisons were based around two metrics: the correlation between the intensity-functions, and the dynamic time warping (DTW) parameter.

The results underscored the conclusion of the previous study: the correlation of the functions is closely related to the human participants' ability to tell the motions apart. The more similar the functions were the harder it was for human observers to distinguish between the videos. The results of the video analysis contribute to the fundamental research

of human perception, by virtue of discovering the basic visual information necessary for the recognition of movements, as well as increasing the accuracy of machine vision, which e.g. can be used to automatically analyze footage from body cameras worn by police officers.

Tartalomjegyzék

1. Bevezetés

1.1. Irodalmi alapok, kutatási cél

1.2. Hipotézisek

2. Módszerek és megoldás

2.1. Nyers videók és azok módosítása

2.2. Frame differenciálás

2.2.1. Bemutatása

2.2.2. Alkalmazásad

2.3. Lineáris interpoláció megvalósítása

2.4. Dynamic Time Warping (DTW)

2.5. Lag-korreláció

3. Eredmények

3.1. Adatok feldolgozása

3.2. Kapcsolat a belső és külső nézetből felvett videók között

3.3. A Masoner kísérlet és az elemzés eredményei

3.4. A két kutatásból kapott eredmények összehasonlítása

4. Tárgyalás (Diszkússzió)

5. Felhasznált irodalom

1. Bevezetés

1.1. Irodalmi alapok, kutatási cél

Az emberi vizuális észlelés alapját a környezetben található fény mintázata szolgáltatja. Ezen mintázat adott fényforrásokból származó fénynyalábok és a különböző felületekről visszaverődött fény egyvelege. A tér minden egyes pontja ezen mintázat egyedülálló kombinációjával írható le. Gibson [1] a tér fényviszonyait környezeti optikai eloszlásnak (ambient optic array) tekintette. A mozgásban lévő test érzékelése ezen eloszlás mintázatának a változásaival jellemezhető, amit optic flow-nak nevezünk. Nyitott kérdés, hogy a mozgásban lévő testek és tárgyak érzékelése az optic flowban rejlő fénymintázatváltozás észlelésén vagy egyéb tulajdonságok (körvonalak, felület textúra, színek, árnyalatok stb.) felismerésén alapszik.

Johansson[2] sikeresen demonstrálta, hogy az észleléshez elegendő az optikai eloszlás változásának követése. A 70-es és 80-as években végzett kísérletsorozatban az emberi test különböző pontjaira (csukló, könyök, térd, fej, stb.) apró fényforrásokat rögzítettek. Fekete háttér előtt fekete ruhába öltözött kísérleti személyekről videó felvételeket készítettek, amelyeken csak apró fénypontok látszottak. Johansson és kollégái bebizonyították, hogy a kísérleti alanyok sikeresen fel tudtak ismerni sokféle emberi tevékenységet (gyaloglás, futás stb.) pusztán a fénypontok mozgásmintázata alapján. A fénymintázatok egzakt elemzése azóta is várat magára és hiányt képez mind a gépi látás, mind a humán észlelési szakirodalomban.

A jelen kutatásunk alapját Masoner [3] kísérletsorozata adta. Ennek a célja annak a bemutatása, hogy az emberek képesek pusztán belső (egocentrikus) nézetből a mozgó ember testét nem látva be azonosítani a mozgást, a videón látható optic flow alapján. Összesen háromféle kísérletet használtak a kutatásban. Az elsőben a résztvevőknek egy videón látható mozgást kellett felismernie és azt szabad formában megnevezni. Erre egy online kérdőív keretein belül került sor. A második kísérlet hasonló volt annyi különbséggel, hogy ebben a válasz megadása során előre megadott mozgásokból lehetett választani. A harmadik kísérletben a résztvevők először egy belső (egocentrikus) nézetből készült videót láttak és ezután kellett eldönteni, hogy kettő külső nézetből készült videó közül melyik mozgás volt látható az első felvételen. A mozgás párok úgy lettek kialakítva, hogy azok azonosítása ne legyen triviális. Ezek a párok a guggolás és felülés, előre haladó szökdelés és kocogás

valamint az egy helyben ugrálás és a jumping jack. Az egocentrikus videó, ahol csupán a test mozgásának a következményei (de maga a test nem) láthatóak a homlokra rögzített kamera mozgása által, és a külső nézetes videó összehasonlítása fontos szempont volt Masoner harmadik kísérletében. Az eredmények azt mutatták ki, hogy a kísérleti alanyok sikeresen tudták beazonosítani ugyanazt a tevékenységet különböző kamera perspektívából. Ez az eredmény azt jelenti, hogy az optic flow invariáns mintázata, ami mindkét videófelvételen jelen van, érzékelhető. Ez az eredmény alátámasztja Johansson eredeti kísérleti eredményeit, ahol a test nem volt látható, de a mozgás ennek ellenére is felismerhető volt.

A jelen dolgozat célja ezen optic flow mintázatának az elemzése Masoner videófelvételei felhasználásával. Azt szeretnénk elérni, hogy a humán észlelés alapjául szolgáló optic flow mintázat leírásához találjunk egy olyan módszert, amivel megjósolható a humán válaszadás alakfelismerési, tevékenység-felismerési és egyéb észlelési feladatok során. Az optic flow megfelelő összehasonlítási módszerének felfedezése segíthet a gépi látásrendszerek fejlesztésében és automatizálásában is. Ezen dolgozat keretein belül mi a Masoner harmadik kísérletének eredményeit hasonlítjuk össze a számítógépes analízissel, mert ebben a kísérletben volt legkevésbé meghatározó a mozgásformák nevének az ismerete, ezáltal a válaszadásban a lingvisztikai tudásból származó hatások nem voltak mérvadóak. A kutatás során a Masoner által végrehajtott kísérlet eredményeit vizsgáltuk és hasonlítottuk össze a saját számítógépes algoritmus alapú eredményeinkkel, ezáltal próbálva egyszerre validálni azokat és közelebb kerülni ahhoz az információhoz, ami egy ember által felhasználható a rendelkezésre álló vizuális adatok közül.

Az ember saját testének mozgását, jelentős részben a látómezejében lévő tárgyak és felületek mozgása alapján ítéli meg. Lishman és Lee [4] kísérletei során egy mozgó falakkal elhatárolt szobában álló emberek mozgását vizsgálták. A kísérleti alanyok tudta nélkül a falakat egy motor segítségével előre és hátra mozgatták úgy, hogy az ne legyen tudatosan észrevehető. A kísérleti alanyok álló helyzetben a tudomásukon kívül a szobával együtt elkezdtek előre-hátra dőlöngélni, másrészt amennyiben a tesztalanyok előre kezdtek sétálni, viszont a szoba tőlük elfele való mozgása gyorsabb volt, mint saját előrehaladásuk, úgy érezték hátrafelé mennek. Rogers [5] ezekből a mérésekből azt a következtetést vonta le, hogy az ember látótere által szolgáltatott információ képes felülmúlni a többi nem vizuális adatforrást, például az ízületekben található kinesztetikus receptorok aktivitását, amelyeknek a feladata a testtartás irányítása. Emellett lényeges még, hogy maga a vizuális információ valamilyen

szinten erősebb és nem feltétlenül memória vagy ismeret alapú, hiszen a tesztben résztvevők tudatában voltak saját előre felé való lépésüknek, mégis ellentétes irányú mozgást érzékeltek.

Az előbbi kutatások egyik problémája, hogy a szoba falai közel vannak az alanyhoz, ezáltal megnövelik a látómezőben való mozgás nagyságát. Ezt egy későbbi kísérlet alatt nagyobb szobában tesztelték és igazolták [6]. Emellett a szobában a talaj nem mozgott, ezáltal nem lehetett a kísérletet reprezentatívnak tekinteni az emberi egyensúlyozás szempontjából. Mindazonáltal a mi esetünkben nem egy egész térrésről, hanem videófelvevételekről van szó, amelyek belső nézetben lényegében csak egy falat tartalmaznak. Hasznos tulajdonság még az is, hogy a fal közelsége által a kamerához viszonyított relatív mozgás elnagyojja a mozdulatok által kiváltott látómezőn belüli változásokat.

1.2. Hipotézisek

Mint azt már említettük, a Masoner kísérlet harmadik tesztjét használtuk kiindulópontként, és 3 hipotézist állítottunk fel. A harmadik kísérletnek a használatából kifolyóan és az eredmények alátámasztásának érdekében feltételezzük, hogy nagyobb fokú az elemzett hasonlóság ugyanazon mozgások két perspektívájú felvételei között, mint különböző belső- és külső kameraállású felvételek esetében. Ez alapján például a guggolás két perspektívájának a hasonlósága nagyobb kellene, hogy legyen, mint a guggolás belső nézetének és az ülés külső nézetének.

A második hipotézis, amit felállítottunk annak a megfogalmazása, hogy a különböző perspektívák közti hasonlósági fok relációt fog mutatni az online kísérlet eredményeivel. Ez magában hordozza, hogy a vizuális információ, ami alapján a kísérlet résztvevői párosították a mozgásokat egyezik az általunk kinyert információval, ez az információ mindkét perspektívában érzékelhetően jelen van és az eredményekből kifolyóan az emberi szem által felfogható. Feltevésünk következménye, hogy a vizuális információ invariáns [7], azaz független a kamera nézőpontjától.

A harmadik és egyben utolsó hipotézisünk munkánk közben alakult ki, méghozzá a különböző hasonlóságot vizsgáló algoritmusok alkalmazásának indoklásakor. Ez a feltételezés a DTW-t [8][9] tekinti a legalkalmasabb hasonlóság vizsgáló algoritmusnak, méghozzá azért, mert a vizuális tér változásainak vizsgálatakor el lehet tekinteni az időbeli lefutás hosszától [10]. A mi esetünkben a jellegzetes karakterisztikák időbeli eltérése két mozgásminta közt a hasonlóság mérőszámának nem reprezentatív értéket tulajdonított, hiszen például a tiszta le-

és felmozgás a guggolás és ülés esetében hasonló azonban időben a kezdéshez képest nem ugyanakkor jelennek meg.

2. Módszerek és megoldás

2.1. Nyers videók és azok módosítása

Masoner harmadik kísérletében szereplő videók, amelyeket feldolgozunk, külső nézetben, illetve rezgéscsillapító nélküli fejre, valamint mellkasra erősített GoPro kamerával készültek. A felvételek 30 képkocka/másodperc (frame rate) és 1920x1080 pixel felbontásúak. A kamera jellegzetessége a széles látószög, ami torzítja a képet és növeli a zavarások számát a videóban (különböző környezeti elemek megjelenése és eltűnése a mozgások során zavaró hatást kelt).

A könnyebb elemezhetőség érdekében a videókat kézzel felvágtuk külön mozdulatokra. Mozdulatok alatt úgymond “mozgásbeli ciklusokat” kell érteni. Tehát például guggolásnál egyenes testtartásból, leguggoláson keresztül, ismételt teljes felegyenesedésig tartó szakaszt értjük, futásnál a kéz előre- és hátra lendítése alapján határoztunk meg egy-egy ciklust, azaz periódust. A kézi videóvágás időigényes folyamat, de gépi vágásnál gondot okozott a pontos és egyben szemléletes (pl. egyenes testtartás) ciklushatárok megkeresése. Szerencsére a Masoner kutatásából átvett videók kézzel viszonylag gyorsan felvághatóak voltak, de nagyobb adatbázison dolgozva mindenképp gépi megoldást kellett volna választanunk.

2.2. Frame differenciálás

2.2.1. Bemutatása

A jelenlegi kutatásban alapvető emberi mozgásokat vizsgálunk, amelyek viszonylag egyszerűek; 2 dimenzióban, azaz mindössze az y (fel és le) és z (előre hátra, ez direkt nem mérhető de kikövetkeztethető) tengelymenti elmozdulások vannak jelen. Néhál x irányú mozgás is megjelenik, de ez nem része a testmozgásnak így elhanyagoltuk.) A videók egyszerűségére tekintettel az elemzés metódusa is egyszerű: A videókat frame differenciálással vizsgáljuk, vagyis két képkocka között elmozdult pixelek számát vesszük figyelembe. Ez azt jelenti, hogy a videó minden képkockáját szürkeárnyalatossá alakítjuk, ezután „kivonjuk” a két képet egymásból. Az 1. és 2. ábrán egy videófelvétel két egymást követő képkockája látható.

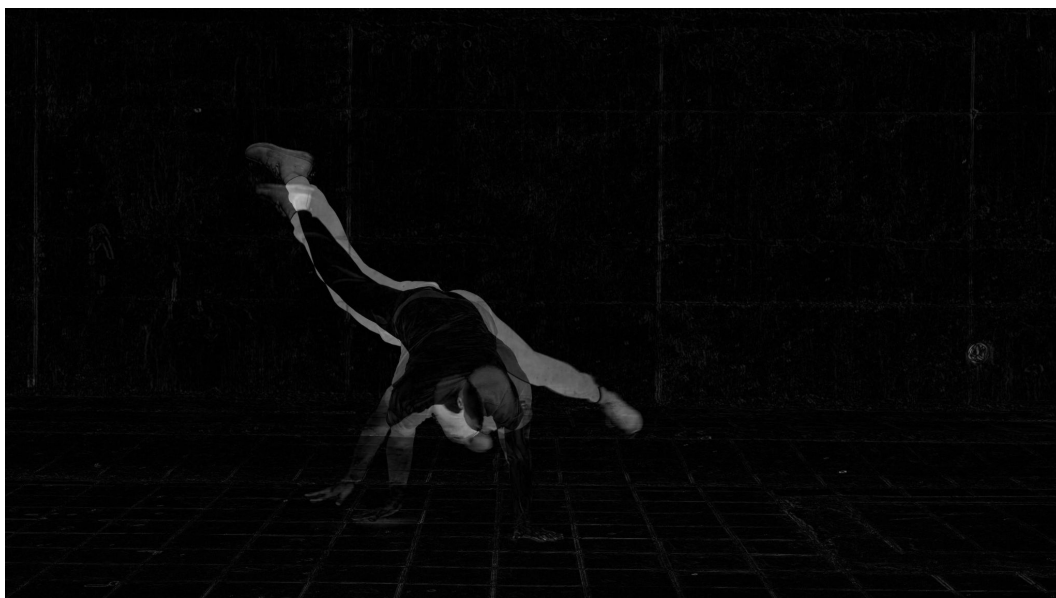


1. Ábra: Az összehasonlítandó videofelvétel szürkeárnyalatossá alakított első képkockája (ez a videó pusztán szemléltető példa, ami nem szolgált ingerként a humán kísérletben, sem a videóanalízisben) [11]



2. Ábra: Az összehasonlítandó videofelvétel szürkeárnyalatossá alakított második képkockája

A két kép felfogható két mátrixként, ahol minden mezőben az adott pixel árnyalata található (digitális kép révén, relatív digitális intenzitásban, DAC-ben mérve), számértékben ez 0-tól 255-ig terjed (255=fehér, 0=fekete). Az összetartozó mezőket vonjuk ki egymásból majd vesszük az eredmény abszolút értékét.



3. Ábra: Az összehasonlítandó képkockák pixelintenzitásának kivonásából készített "differenciált" képkocka

Az így kapott mátrix megmutatja melyik pixelek színintenzitása változott meg a két képkocka között: ha egy mezőben 0 található akkor annak színe nem változott, bármilyen más érték esetén történt valamilyen változás (lásd a 3. Ábrát). Ez a mátrix is megfeleltethető egy szürkeárnyaltos képnek. Ezt a képet binarizáljuk az alábbi módon: minden pixelt, amelynek értéke nem 0 azaz nem teljesen fekete azt fehérre változtatjuk (lásd a 4. Ábrát). Ezzel azt érjük el, hogy minden pixel, ahol történt változás fehér színűre fog változni.



4. Ábra: Binarizált "differenciált" képkocka

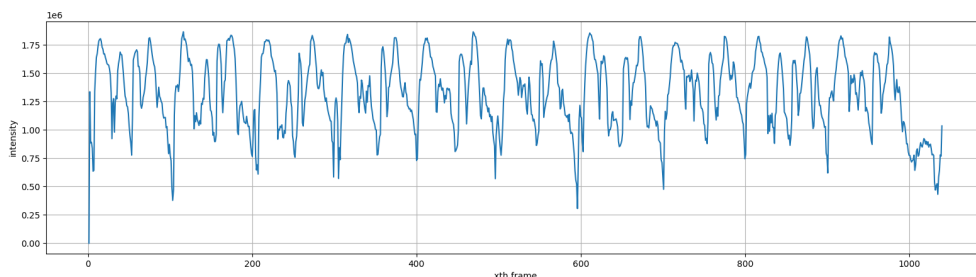
A binarizált képkockán lévő fehér képpontokat összeszámoljuk - ez egyfajta leírást ad a videó mozgalmasságáról, erre a továbbiakban intenzitásként fogunk hivatkozni. Itt fontos megjegyezni, hogy jelenleg nem vizsgáljuk a mozgások irányát, a pixelek elmozdulásához nem rendelünk vektorokat, az algoritmus kizárólag a mozgó pixelek számát figyeli, azok irányát nem. Így ez az algoritmus nem alkalmas önmagában mozgások beazonosítására. Az irányok elhanyagolásával a videókban lévő információ jelentős részét figyelmen kívül hagyjuk. Azonban Masoner harmadik kísérletében szereplő párok (guggolás - felülés, szökdelés - kocogás, ugrálás - jumping jack) sok hasonlósága között vélhetően a mozgás iránya hasonlít legjobban.

A kutatásunk során főként mi is ezeket a párokat vizsgáljuk. A közös irányt elhanyagolva kevesebb és könnyebben feldolgozható adatunk marad hátra, a párok közti mintázatbeli különbségeket egyszerűbben meg tudjuk határozni. Szélesebb körű felhasználásra, más jellegű vizsgálatokra viszont célszerű lehet eltérő, összetettebb algoritmusokat alkalmazni, ezek közül pár opcióra később ki fogunk térni.

2.2.2. Alkalmazása

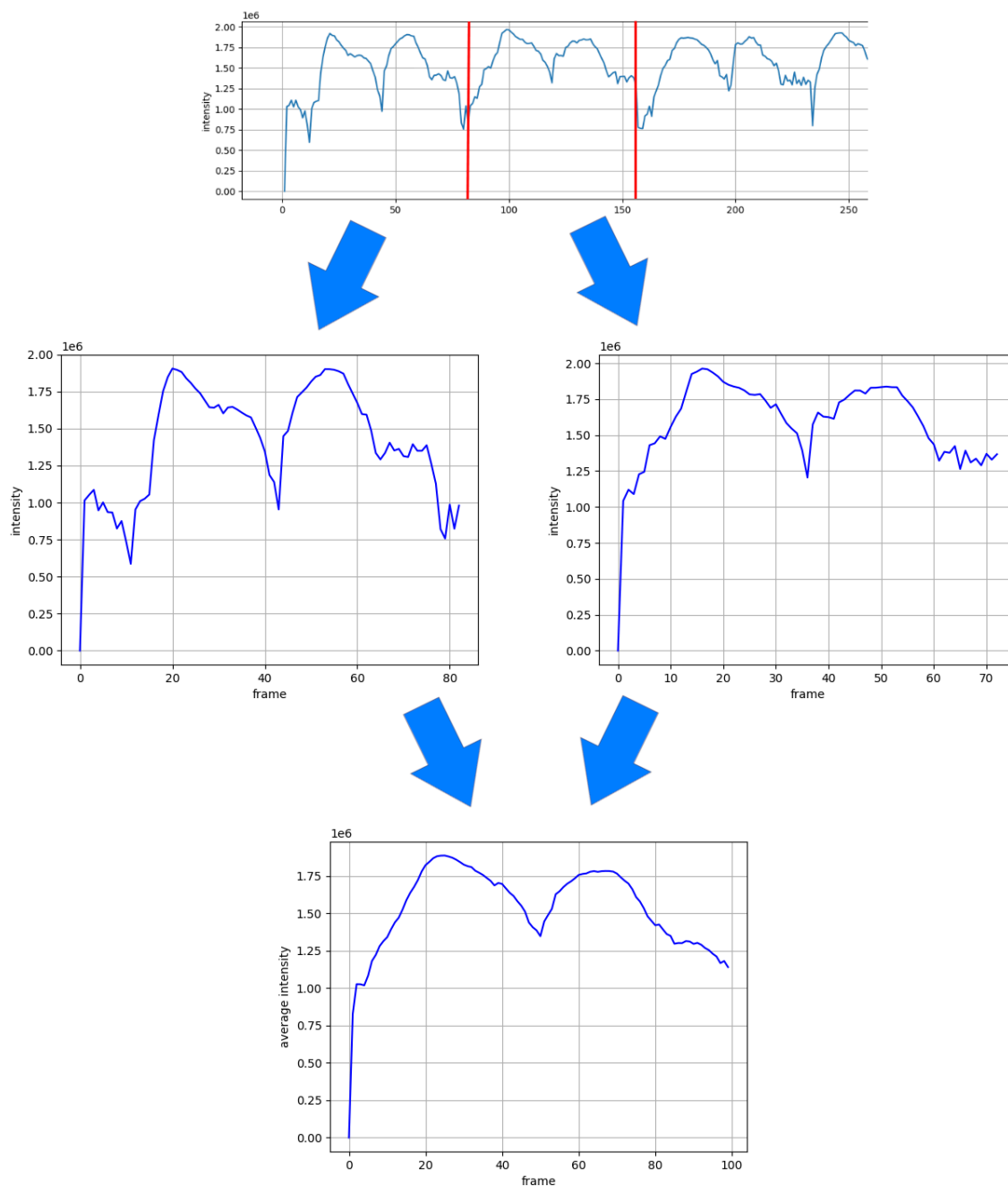
Kezdetben minden mozgásvideó minden képkockáján futtattuk a bemutatott frame differenciálás algoritmust majd ábrázoltuk az intenzitást az idő függvényében (az első időpillanat az első képkockának felel meg). Az intenzitás értékeket érdemes normalizálni felbontás szerint ($\text{intenzitás}/(\text{szélesség} \cdot \text{magasság})$) de jelen kutatásban minden felvétel ugyanazzal a kamerával készült így ez nem befolyásolja az eredményeket.

A videónként kapott intenzitás-idő függvényeken jól látszanak a mozgások karakterisztikái, azonban a felvételen szereplő zaj, jelentéktelen apró eltérések és anomáliák is részei a függvényeknek.



5. Ábra: A felülés belső (egocentrikus) nézetből (teljes mozgatlansor)

Az anomáliákat legegyszerűbben átlagolással tudjuk kiszűrni, elsimítani. Ehhez szükséges a függvényünket, ezzel a videókat is darabokra, mozdulatokra vágni, ezeket fedésbe hozni, majd adatpontonként átlagolni. Ezt a folyamatot ábrázoljuk a 6. Ábrán.

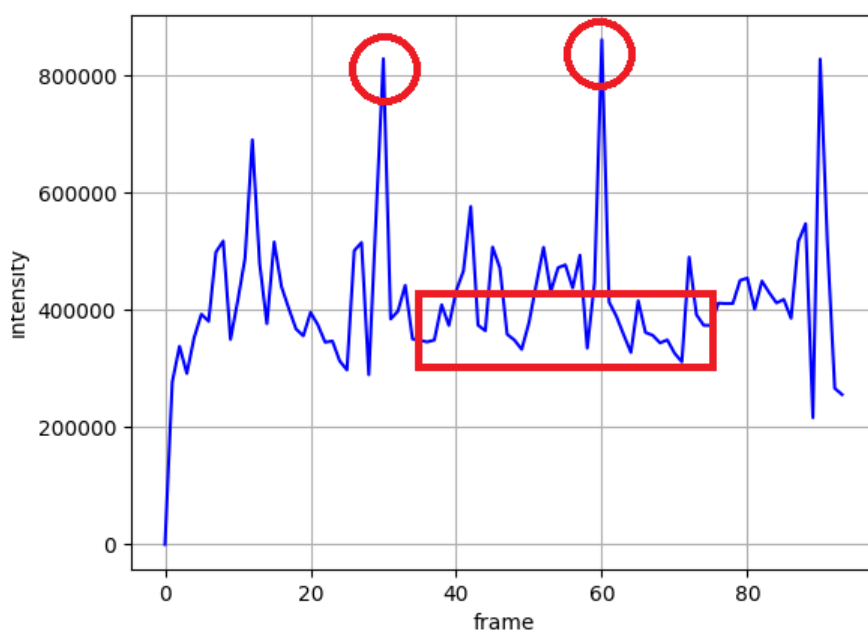


6. Ábra: Mozdulatsor periódusokra (ciklusokra) való felbontása (felső grafikon), a kivágott ciklusok szemléltetése (a példa kedvéért csak 2, középső két grafikon) és a ciklusok átlaga, vagyis az átlagos mozdulat (alsó grafikon)

A legtöbb intenzitásfüggvény jellegre hasonló: ciklusonként egy vagy kettő jelentős lokális maximum, illetve minimum található bennük. Kezdetben maximum és minimum helyek illesztésével próbáltuk fedésbe hozni a függvényeket az átlagoláshoz.

Minimumok illesztése:

A legtöbb mozgás esetén nehéz egyértelműen megtalálni a függvény abszolút minimumát, mivel a zavarások miatt több lokális minimum is körülveszi, melyek értékeinek kicsi az eltérése az abszolút minimumhoz képest. Emiatt a ciklusokra vágás pontatlan, az így fedésbe hozott függvények között viszonylag nagy eltolódások lehetnek, ezzel egy “elmosódott” átlagfüggvényt kapunk.

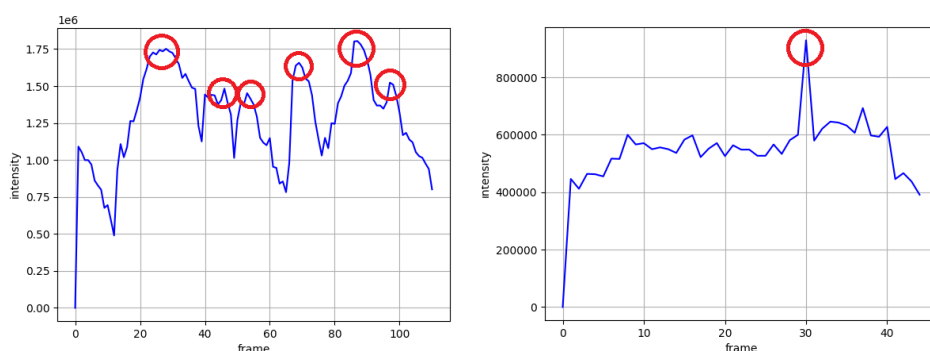


7. Ábra: Minimumkeresés problémája

Maximumok illesztése:

A minimumokkal ellentétben a függvények maximum helyei egyértelműbben meghatározhatóak. Meghatároztuk a függvények abszolút maximumát illetve a második legnagyobb értékű lokális maximumot (ábra a két csúcsról), és képeztük a maximum helyek átlagát, majd ez alapján illesztettük a függvényeket (a két csúcs közti tartományok azonos hosszúságúra nyújtásával lehetne a módszert még pontosítani). Ez a módszer azonban csak

olyan videókon működött jól amelyeknek legalább 2 domináns lokális maximuma volt, így ezt a módszert is el kellett vetnünk.



8. Ábra: Maximumok számának változása

Ezek után választottuk a harmadik és végső módszert, a videók ciklusokra vágását kézzel, amit a fejezet bevezetésében is említettünk. Gondot okozott, hogy az egyes mozdulatok különböző ideig tartottak (nem csak különböző, de azonos mozgássorozatok esetén is). Ezért idő szerint normalizálnunk kellett a grafikonokat. Módszernek a lineáris interpolációt [12] választottuk.

2.3. Lineáris interpoláció megvalósítása

Mindegyik mozgásciklust 100 adatpont hosszúságúra alakítottuk. Az eredeti adatsor (függvény) két végpontját „rögzítettük” (az első és az utolsó adat a normalizált adatsor első és utolsó adata). Ezután pivot pontokat hoztunk létre (ezek az új skála helyeit jelentik), és átskáláztuk az eredeti függvényt a kiválasztott 100-as skálára. Az egyszerűség kedvéért lineáris közelítést alkalmaztunk az új intenzitás értékek meghatározásához. Az immáron idő tengely irányában azonos hosszúságú grafikonokat átlagoltuk. Az átlagolással a mozdulatok közti eltérések, zaj, anomáliák hatását csökkentettük, és csak a mozdulatokat jellemző karakterisztikus mintázat maradt meg (ábra), tehát megkaptuk a mozdulatok átlagos intenzitás függvényét.

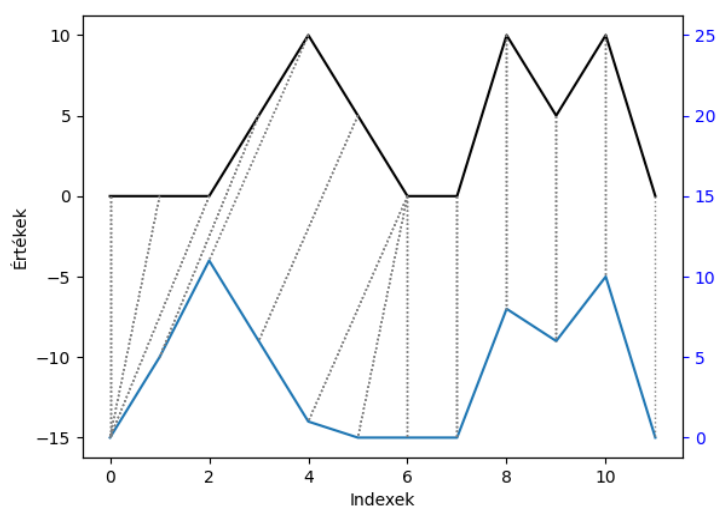
További cél volt a mozgások hasonlóságának mérése és számszerűsítése. Ehhez két külön algoritmust használtunk melyek a DTW [8][9] és a Lag-korreláció (Lag correlation, Cross correlation, keresztkorreláció). A Dynamic Time Warping két idősor hasonlóságának mérésére alkalmas algoritmus, a két bemeneti idősor között egy átvitt értelemben vett távolságot számol. A módszerről nagyon jó leírások és vizualizációk léteznek (pl.

<https://rtavenar.github.io/blog/dtw.html#dynamic-time-warping>), de a teljesség érdekében mi is bemutatjuk a fontosabb tulajdonságait.

Mint említettük, a mozgások összehasonlításának alap nehézsége, hogy gondosan megvágott videóknál is a mozdulatok egyes részei nem pontosan ugyanakkor történnek, mivel az alany mindig nagyon picit máshogy végzi el a mozgást. Ha például az első ciklusban, tehát az első videón az alany az n -dik képkockánál egyenesedik ki, attól még a következő ciklusban ez valószínűleg a párral előtti vagy későbbi képkockán fog megtörténni. Emiatt nem tudunk egyszerűen képkockáról képkockára hasonlítani.

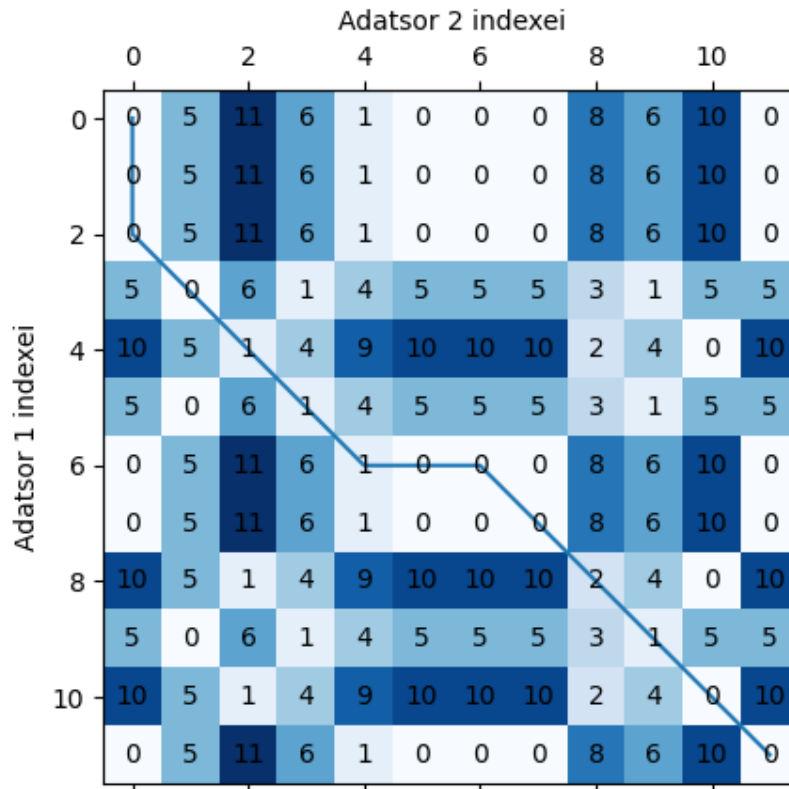
2.4. Dynamic Time Warping (DTW)

A DTW lokálisan “gyorsítja vagy lassítja” az időt, a gyakorlatban az egyik idősorból egy pontot a másiktól több ponttal társít (ábra), ezzel az adatsor többi része jobb fedésbe kerül. Természetesen a nyújtás mértéke mintegy szorzóként hat a végső távolságra.



9. Ábra: A DTW módszer bemutatása tetszőleges idősorok példáján.

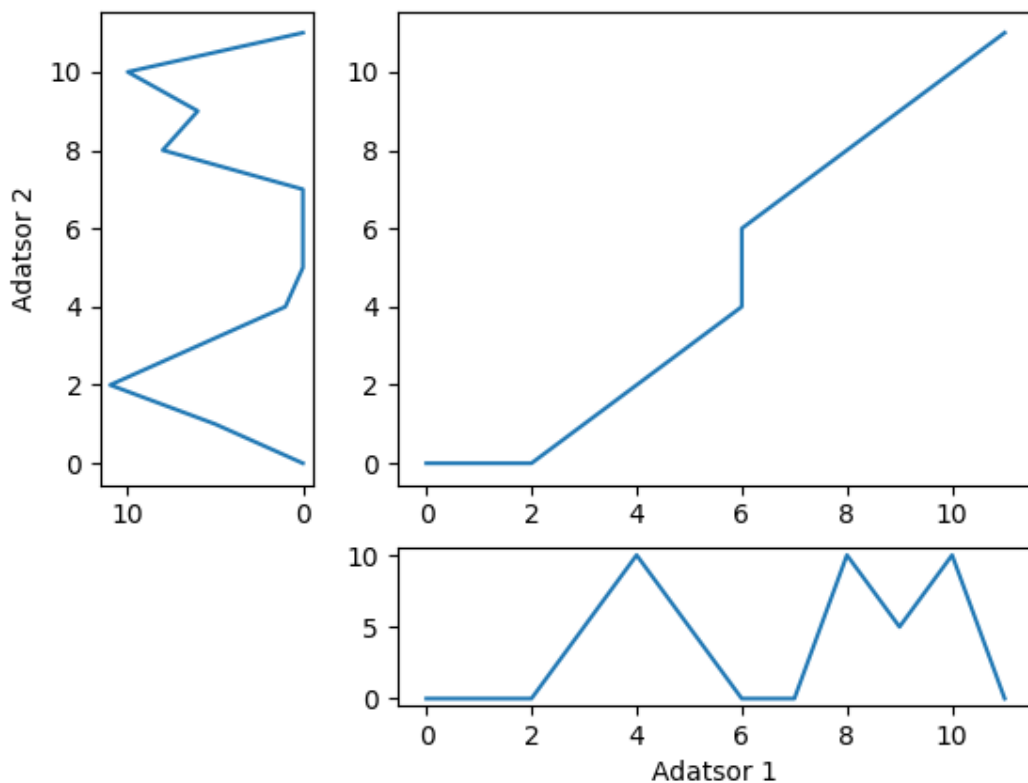
Ezek alapján mondhatjuk, hogy a DTW megtalálja az adatsor jellegre hasonló, de időben összenyomottabb, illetve széthúzottabb részeit, ami megfelel annak, hogy egy mozdulatot az alany kicsit gyorsabban vagy lassabban, több vagy kevesebb szünettel végez el, és ezekre viszonylag kis távolságot ad. Térjünk ki arra is, hogy az algoritmus hogyan számolja ki két adatsor között a “távolságot”, vagyis a DTW értéket. Ez jól szemléltethető az alábbi módon: Képzeljünk el egy mátrixot annyi sorral, amilyen hosszú az egyik adatsorunk, annyi oszloppal, amilyen a másik (ábra). Minden mezőbe írjuk be, hogy az adatsorok megfelelő indexű (sorszámú) elemei között mekkora a különbség abszolút értéke (távolság).



10. Ábra: A DTW-érték számításának bemutatása tetszőleges idősorok példáján, mátrix ábrázolás segítségével

Ebben a mátrixban “kiválaszthatunk” mezőket, ez megfelel annak, hogy két pontot összekötünk, a mezőben szereplő érték hozzáadódik az össztávolsághoz (ami végül a DTW értékünk lesz). El kell jutnunk a bal felső sarokból a jobb alsóba mezők kiválasztásával úgy, hogy csak jobbra, lefele, vagy átlósan jobbra lefelé léphetünk. Ez pont megfelel annak, hogy a pontokat egymáshoz társítjuk időbeli megcserélése nélkül. A bejárt út alatt lévő számokat összeadva megkapjuk az össztávolságot, és a legkisebb elérhető össztávolság a DTW érték.

Az időnyújtást és az út bejárását még egy harmadik módon is szemléltethetjük (ábra). Az ábra jobb felső diagramján a $[0;0]$ pontból a kék vonal mentén kell haladnunk. Ha a vonal vízszintesen megy, akkor a vízszintes tengelyhez tartozó adatsoron haladunk, a másik adatsoron várakozunk. Ha függőlegesen, akkor fordítva. Ha pedig átlósan, akkor mindkét adatsoron haladunk.



11. Ábra: A DTW algoritmus bemutatása tetszőleges idősorok példáján grafikonos ábrázolás segítségével

A DTW értéknek van egy normalizált változata is, ahol az eredményül kapott értéket, távolságot elosztjuk a két adatsor hosszának összegével. Ezzel különböző hosszúságú adatsorokon számolt DTW értékeket lehet összehasonlítani, de számunkra csak a kisebb számokkal való dolgozásért volt előnyös a normalizálás. A továbbiakban mi ezzel, a normalizált DTW értékkel dolgozunk.

2.5. Lag-korreláció

A lag-korreláció (lag correlation), vagy másnéven keresztkorreláció (cross correlation) szintén idősorok hasonlóságának mérésére alkalmas algoritmus. Az idősorok között különböző csúsztatásokat (lag) vizsgál, az elcsúsztatott idősorok között Pearson-féle korrelációt számol.

Erről a módszerről is jó összefoglalások léteznek (pl. [13]), de röviden mi is bemutatjuk.

A módszernél meg kell adni egy maximális csúsztatás értéket, amennyivel odébb tolhatja a két adatsort. Mi a 100 hosszú adatsorainkat 30, 20, 10, 5 és az ellenőrzés kedvéért 0 maximális csúsztatással hasonlítottuk. A módszer nagy hátránya, hogy az idősorokat

egészként csúsztatja, tehát ha az alany az adatsor, vagyis videó közepén egy pár képkockával többet időz pl. a felkeléssel, mint egy másik alkalommal, akkor a két adatsort semmilyen eltolással nem lehet megfelelően egymásba csúsztatni. Minél több ilyen apró eltérés van a felvételen, annál nehezebb fedésbe hozni az adatokat, még ha a mozdulatok hasonlóak is. Az adatsoraink viszonylag rövidek, nagyjából 100 képkockából, így adatpontból állnak. Emiatt vélhetően nem tud sok apró eltérés felhalmozódni, ami használhatatlanná tenné a lag-korrelációt, így a DTW mellett ezt a módszert is megéri, mint egy alternatív megoldást kipróbálni. Azonban az elméleti szinten is jelenlévő hiányosságokat a kezdeti számításaink eredményeinek nehezen kivehetősége is alátámasztotta, így végül a lag-korreláció alapú analízist elvetettük. Tehát a DTW-re, mint egy elméleti és gyakorlati szinten is jobban a feladatunkhoz illő algoritmusra hagyatkoztunk.

3. Eredmények

3.1. Adatok feldolgozása

Ahhoz, hogy értelmezhetőek legyenek a DTW algoritmussal kapott eredmények, táblázatba foglaltuk a kapott adatokat. A DTW algoritmust periódusonként (vagyis a vágott videókon) futtatuk le. A legkisebb ismétlésszám vagyis ciklusszám a mozgásainkban négy ismétlés volt (az egyik szökdelés videó csak négy teljes szökdelés-ciklusra volt felvágható). A lent látható ábrán (ábra hivatkozás) még az ismétlésszámra való tekintet nélkül szerepelnek az adott összehasonlítás DTW értékei (vagyis az összes ciklusból számolt átlagfüggvények összehasonlításait tartalmazza az ábra), azonban a statisztikai módszerek alkalmazhatósága érdekében a később elemzett eredményeink esetében minden mozgásból csak az első négy ciklust átlagoltuk és hasonlítottuk a többi mozgás első négy ciklusából vett átlaghoz.

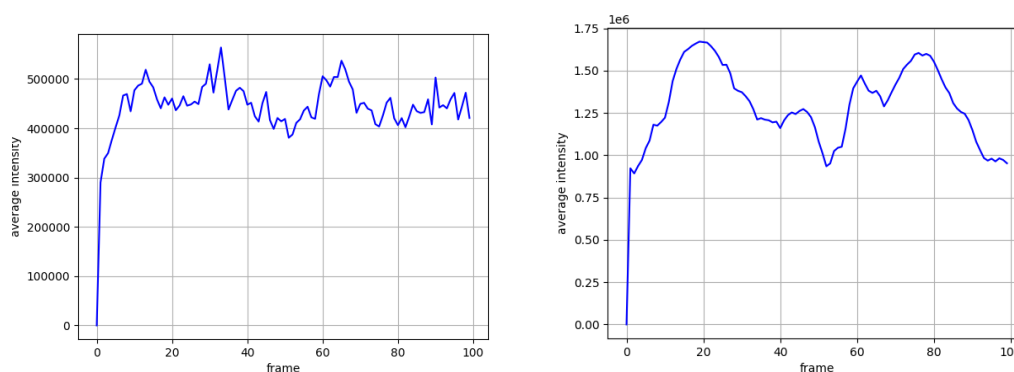
	CfFirst_Lunge	CFirstJacks	CFirstJog	CFirstSit	CFirstSkip	CFirstSquats	CFirstStep	CFirst_Jump	HFirstJacks	HFirstJog	HFirstJump	HFirstLunge	HFirstSit	HFirstSkip	HFirstSquat	HFirstStep	ThirdJacks	ThirdJog	ThirdJump	ThirdLatStep	ThirdLunge	ThirdSit	ThirdSkip	ThirdSquat
CfFirst_Lunge	0.0	0.83	0.53	0.58	0.53	0.36	1.09	0.64	0.71	0.82	0.69	0.4	0.94	0.7	0.84	0.85	5.39	5.4	6.39	6.26	6.98	6.86	5.19	6.67
CFirstJacks	0.83	0.0	0.72	1.19	0.68	0.67	2.27	0.41	0.43	1.27	0.45	0.75	2.1	0.97	1.34	2.06	5.67	5.42	6.27	5.83	6.38	6.31	5.08	6.2
CFirstJog	0.53	0.72	0.0	1.57	0.2	0.67	2.54	0.56	0.73	1.78	0.67	0.63	1.95	1.23	1.5	2.4	5.67	5.3	6.54	6.07	6.25	6.23	5.06	6.27
CFirstSit	0.58	1.19	1.57	0.0	1.63	0.7	0.65	1.25	0.92	0.78	0.83	0.77	0.64	0.74	0.6	0.59	4.49	4.67	5.5	5.3	6.06	6.01	4.54	5.97
CFirstSkip	0.53	0.68	0.2	1.63	0.0	0.79	2.78	0.64	0.75	1.81	0.72	0.82	1.98	1.16	1.67	2.65	6.38	5.9	6.76	6.51	6.35	6.32	5.65	6.4
CFirstSquats	0.36	0.67	0.67	0.7	0.79	0.0	1.3	0.36	0.63	0.96	0.3	0.21	1.42	0.93	0.62	1.04	5.76	5.65	6.75	6.55	7.27	7.15	5.43	6.94
CFirstStep	1.09	2.27	2.54	0.65	2.78	1.3	0.0	2.14	2.07	0.75	1.71	1.05	0.99	1.2	1.01	0.19	2.57	2.73	3.29	3.07	3.64	3.55	2.38	3.33
CFirst_Jump	0.64	0.41	0.56	1.25	0.64	0.36	2.14	0.0	0.69	1.38	0.26	0.45	1.98	1.11	1.06	1.84	5.76	5.66	6.35	5.93	6.43	6.37	5.21	6.23
HFirstJacks	0.71	0.43	0.73	0.92	0.75	0.63	2.07	0.69	0.0	1.13	0.71	0.86	1.46	0.82	1.12	1.71	5.32	5.01	6.44	5.9	6.62	6.53	4.77	6.35
HFirstJog	0.82	1.27	1.78	0.78	1.81	0.96	0.75	1.38	1.13	0.0	1.42	1.09	1.24	0.25	1.45	0.77	4.24	3.86	4.9	4.59	4.53	4.52	3.61	4.47
HFirstJump	0.69	0.45	0.67	0.83	0.72	0.3	1.71	0.26	0.71	1.42	0.0	0.37	1.51	1.1	0.68	1.54	5.86	5.6	6.49	6.19	6.09	6.03	5.36	5.99
HFirstLunge	0.4	0.75	0.63	0.77	0.82	0.21	1.05	0.45	0.86	1.09	0.37	0.0	1.19	1.16	0.5	0.86	4.58	4.5	5.59	5.36	6.09	5.98	4.27	5.76
HFirstSit	0.94	2.1	1.95	0.64	1.98	1.42	0.99	1.98	1.46	1.24	1.51	1.19	0.0	1.26	0.88	0.84	2.07	2.38	2.73	2.56	3.08	2.98	2.31	3.07
HfirstSkip	0.7	0.97	1.23	0.74	1.16	0.93	1.2	1.11	0.82	0.25	1.1	1.16	1.26	0.0	1.42	1.14	4.78	4.37	5.56	5.28	5.16	5.11	4.14	5.14
HFirstSquat	0.84	1.34	1.5	0.6	1.67	0.62	1.01	1.06	1.12	1.45	0.68	0.5	0.88	1.42	0.0	0.8	3.27	3.55	4.07	3.88	4.64	4.59	3.45	4.45
HFirstStep	0.85	2.06	2.4	0.59	2.65	1.04	0.19	1.84	1.71	0.77	1.54	0.86	0.84	1.14	0.8	0.0	2.77	3.01	3.61	3.48	4.14	4.04	2.73	3.83
ThirdJacks	5.39	5.67	5.67	4.49	6.38	5.76	2.57	5.76	5.32	4.24	5.86	4.58	2.07	4.78	3.27	2.77	0.0	0.12	0.2	0.15	0.45	0.48	0.15	0.46
ThirdJog	5.4	5.42	5.3	4.67	5.9	5.65	2.73	5.56	5.01	3.86	5.6	4.5	2.38	4.37	3.55	3.01	0.12	0.0	0.38	0.31	0.75	0.72	0.11	0.54
ThirdJump	6.39	6.27	6.54	5.5	6.76	6.75	3.29	6.35	6.44	4.9	6.49	5.59	2.73	5.56	4.07	3.61	0.2	0.38	0.0	0.1	0.15	0.17	0.34	0.16
ThirdLatStep	6.26	5.83	6.07	5.3	6.51	6.55	3.07	5.93	5.9	4.59	6.19	5.36	2.56	5.28	3.88	3.48	0.15	0.31	0.1	0.0	0.16	0.19	0.28	0.19
ThirdLunge	6.98	6.38	6.25	6.06	6.35	7.27	3.64	6.43	6.62	4.53	6.09	6.09	3.08	5.16	4.64	4.14	0.45	0.75	0.15	0.16	0.0	0.12	0.64	0.08
ThirdSit	6.86	6.31	6.23	6.01	6.32	7.15	3.55	6.37	6.53	4.52	6.03	5.98	2.98	5.11	4.59	4.04	0.48	0.72	0.17	0.19	0.12	0.0	0.64	0.12
ThirdSkip	5.19	5.08	5.06	4.54	5.65	5.43	2.38	5.21	4.77	3.61	5.36	4.27	3.14	4.14	3.45	2.73	0.15	0.11	0.34	0.28	0.64	0.64	0.0	0.49
ThirdSquat	6.67	6.2	6.27	5.97	6.4	6.94	3.33	6.23	6.35	4.47	5.99	5.76	3.07	5.14	4.45	3.83	0.46	0.54	0.16	0.19	0.08	0.12	0.49	0.0

12. Ábra: Az összehasonlítások DTW értékei (az ábra tartalmazza a lateral lunge és step (oldalra kitörés és lépés) mozdulatok hasonlítását is, viszont ezeket Masoner nem használta fel a harmadik kísérletében, így a későbbi számításainkban sem szerepel.)

A Third és a First kifejezések rendre a külső és belső nézetre utalnak. A C (chest) jelölés a mellkason, a H (head) a fejen elhelyezett kamerára utal. A táblázatban az oszlopban elhelyezett mozgásokat hasonlítjuk össze a sorokban elhelyezettekkel, a cellákban a pár összehasonlításával kapott DTW érték szerepel, 10^{-5} szorzóval skálázva, a könnyebb megjeleníthetőség érdekében. A cella annál sötétebb minél nagyobb a különbség a videók között. A főatlóban 0-k szerepelnek, hiszen a mozdulatokat saját magukkal hasonlítva 0 különbséget kapunk. Az egocetrikus videók közül statisztikai kiértékelésre a mellkasra helyezett kamerafelvételket használtuk fel, mivel azok stabilabb felvételt eredményeztek.

3.2. Kapcsolat a belső és külső nézetből felvett videók között

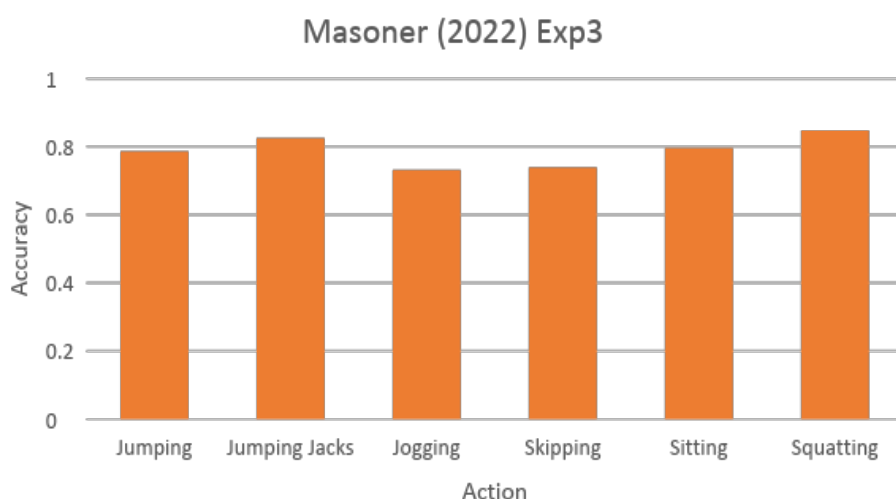
Ezentúl az is látszik, hogy ha különböző nézetből hasonlítjuk össze a videókat, akkor nagyban eltérő értékeket kapunk. Ez magyarázza a bal alsó sarokban és jobb felső sarokban található sötét téglalapot az XX ábrán. Az algoritmusunk felépítéséből adódik, hogy a külső nézetes videókhoz tartozó intenzitásfüggvények átlagos értéke és az intenzitás változása is jóval kisebb mint a belső nézetes videókhoz tartozó, hiszen külső nézet esetén a környezet állónak tekinthető és csak egy mozgást végző személy pontjai lesznek detektálva, míg a másik esetben a környezet minden pontja mozgásban van. Például a [13]. ábrán látható mozdulatnál nagyjából háromszoros intenzitásokat és intenzitásváltozásokat kapunk belső nézetnél, mint külsőnél.



13. Ábra: Bal oldalon: Az ülés külső nézetből vett intenzitásfüggvénye, Jobb oldalon: Az ülés belső (chest) nézetből vett intenzitásfüggvénye

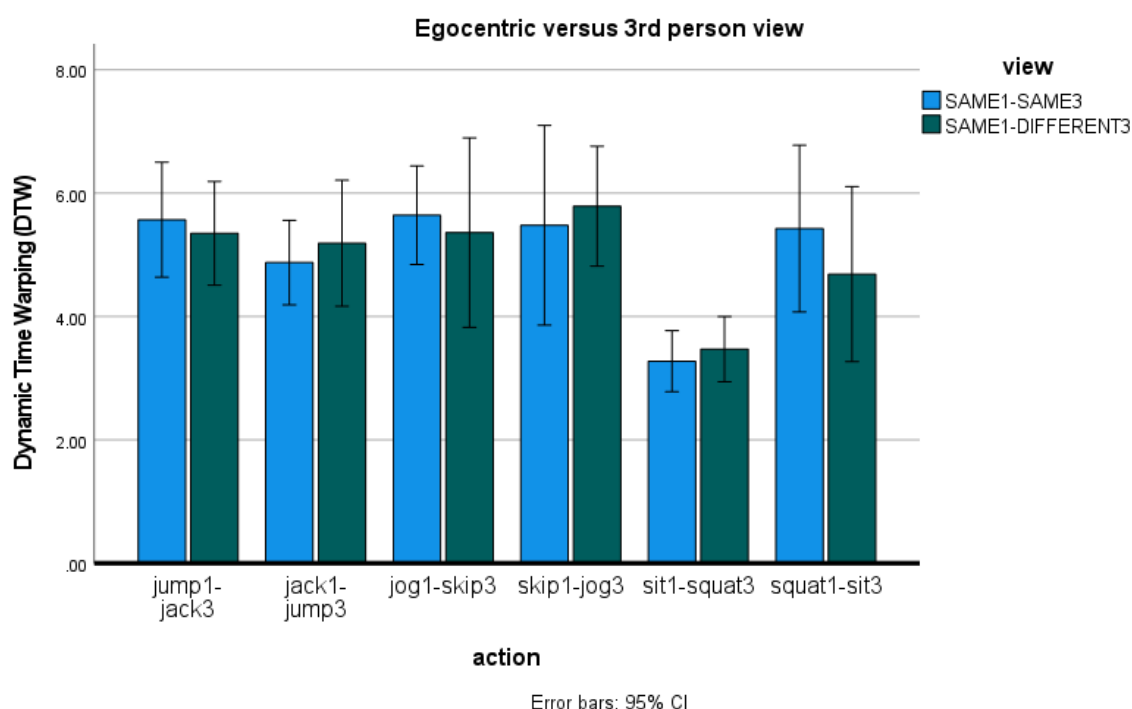
3.3. A Masoner kísérlet és az elemzés eredményei

Ahhoz, hogy jobban megértsük a kapott eredményeket egy rövid összefoglalásra lesz szükség a Masoner kísérlet eredményeiről. A bevezetésben már meg lett említve, hogy a harmadik kísérlet eredményeit fogjuk felhasználni, ahol a videók mellkasra erősített kamerával, vagyis belső nézetből lettek felvéve és prezentálva az online kérdőívet kitöltő tesztalanyoknak.



14. Ábra: Masoner (2022) 3. kísérletének eredményei. A válaszadás pontossága (Accuracy) a különböző mozgások függvényében

Az ZZ Ábrán a narancssárga oszlopok jellemzik Masoner harmadik kísérletének eredményeit. Az ábra y-tengelyén a nulla azt jelenti, hogy senki sem párosította jól a mozgásokat míg az 1 azt, hogy minden párosítás jó volt. Ezt figyelembe véve látszik, hogy minden párosítás tartalmaz nézetek közti invariáns információt a videóban, ami alapján a mozgások párosíthatóak, tekintve, hogy az összes eredmény 0,5 felett van vagyis a véletlenszerű statisztikai eséllynél szignifikánsan nagyobb.



15. Ábra: A DTW által kiszámított hasonlóság a perspektíva (1=egocentrikus; 3=külső) és a mozgástípus függvényében. A DTW változó definíciója alapján nullához közeli értékek nagyobb mértékű hasonlóságot jelentenek.

A CC. Ábrán az egocentrikus nézeteket hasonlítjuk össze a külső nézetekkel az említett mozgáspáronként. Ezzel tulajdonképpen azt vizsgáljuk, hogyan viszonyul egy egocentrikus videó ugyanannak a mozgásnak a külső nézetből felvett videójához (világoskék oszlop), és a vele párosított mozgásnak a külső nézetes felvételéhez (sötétkék oszlop).

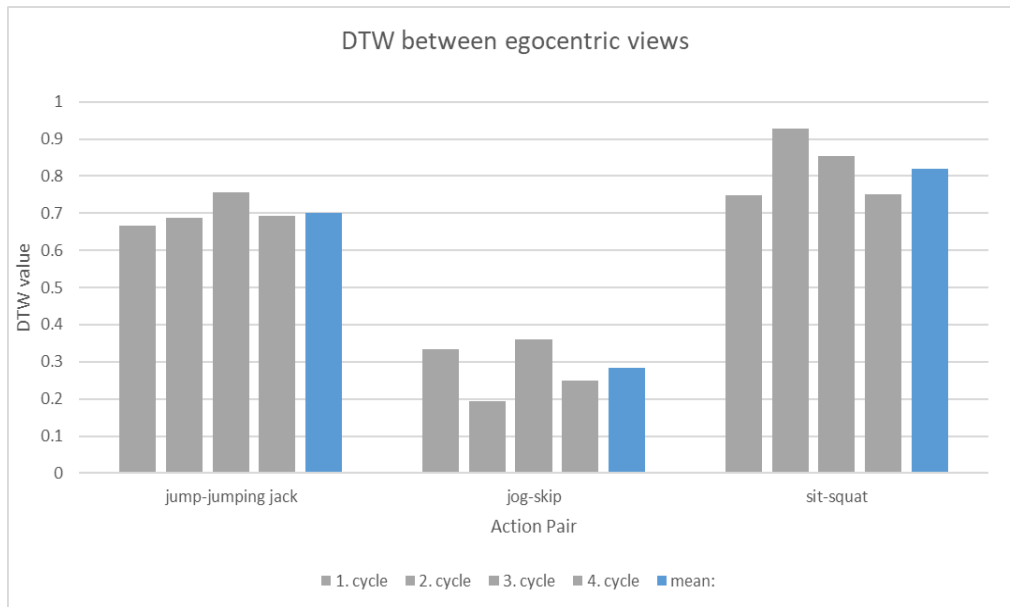
3.4. A két kutatásból kapott eredmények összehasonlítása

Az első és talán legszembetűnőbb, hogy humán kísérletben résztvevő alanyok a felülés és a guggolás mozgásokat tudták a legnagyobb pontossággal megkülönböztetni. Ez megegyezik az elemzésből kapott eredményekkel, abból a szempontból, hogy a (ref) grafikonon a felülés jobban hasonlít a saját külső nézetből felvett videóra, mint bármely másik külső és belső nézet párosítás esetében. Ellentmondást lehet azonban felfedezni abban, hogy bár a felülés (sit) jobban hasonlít a saját külső nézetes videójára addig a guggolásról (squat) készült egocentrikus videó szintén a felülés másik nézetből felvett mozgásmintázatára hasonlít.

Ahhoz, hogy ezt a jelenséget megértsük, szükségünk van egy új tényező bevezetésére, pontosabban átvételére a humán kísérlet paraméterei közül. Az online kérdőív minden kérdés esetében először megmutatott egy videót, ami lehetett akár egocentrikus, akár külső nézetű. Ezután az előbbieken nem megjelenített perspektívából mutattak meg két videót a pár egyik és másik tagjáról.

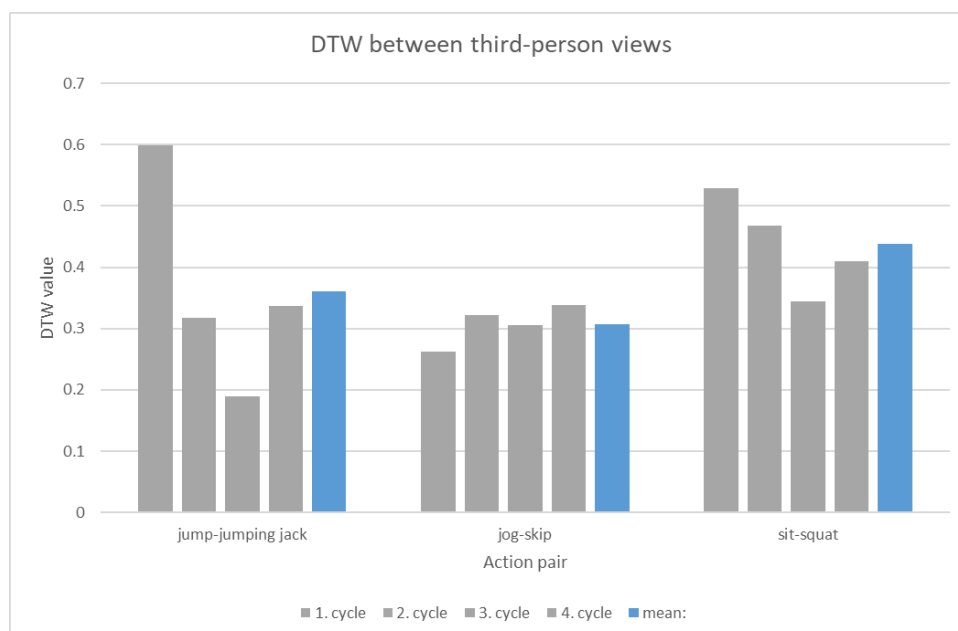
Ez alapján feltételezhető, hogy nem csak az határozta meg a teszt résztvevőinek döntését, hogy mennyire hasonlított egy adott videó a másik perspektívából felvett változatára hanem, hogy a pár két tagja közül melyik hasonlít jobban a másik perspektívára. Így találtunk magyarázatot az előbbi ellentmondásra, ami (ref 1 diagram)-on is látszik, hiszen a felülés belső nézete jobban hasonlít a felülés külső nézetére, mint a guggolás a felülés külső nézetére.

Ez azonban a többi mozgásnál nem érvényesül (kivéve a jumping jacket), éppen ezért elkülöníthető a többi pártól. Viszont van még egy tényező, ami segíti a választásban a megfigyelőket, és ezáltal növeli vagy csökkenti a pontosságot. Mivel a sit-sit és sit-squat belső- és külső nézetű videóján kívül minden más párosítás relatíve nagy DTW értékkel rendelkezik, ezért célszerűbb lehet a két párhuzamosan prezentált videó között keresni a különbséget.



16. Ábra: a DTW által kiszámított hasonlóság egocentrikus videók között

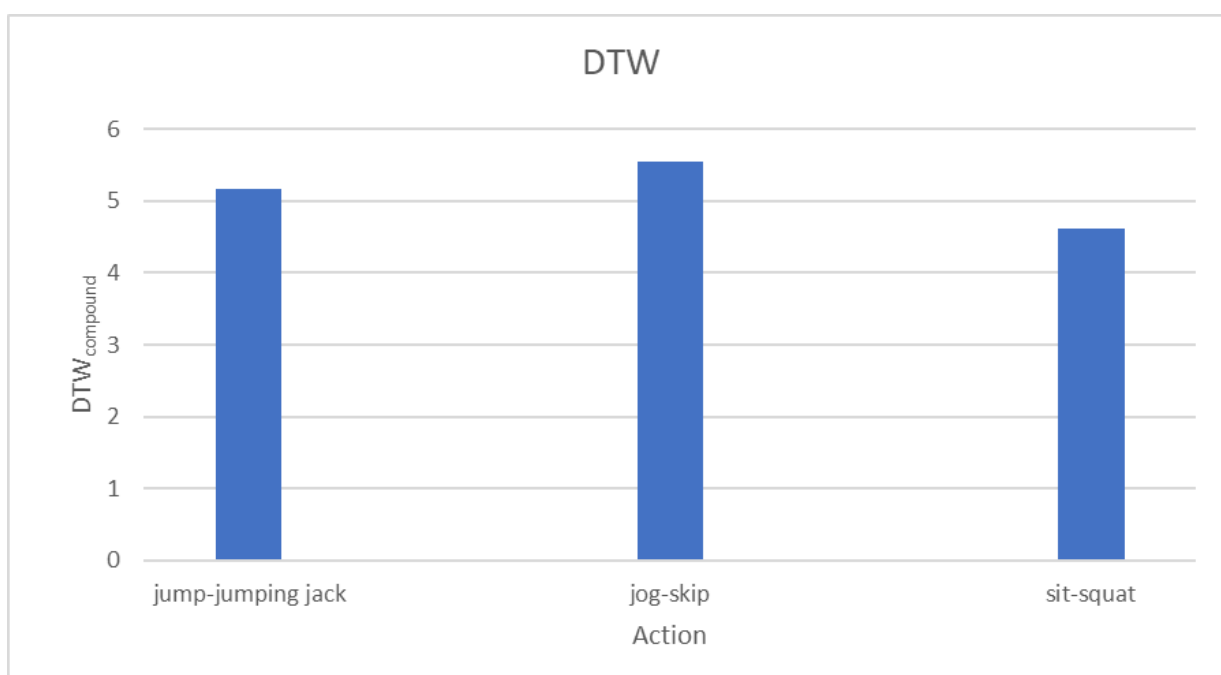
A fenti diagramon a mozgáspárok mindkét tagjára (egocentrikus perspektívából) számított DTW értékek láthatóak. Ez az ábra számunkra egyértelmű magyarázatot tud adni a résztvevők által elért eredményekre, ugyanis míg a külső nézetre esetlegesen mindkét egocentrikus videó közel teljes mértékben hasonlít addig a sit-squat és a jump-jumping jack párosítások egocentrikus videói egymástól jelentősen eltérnek, így már csak a két videó közötti választás is esetlegesen elég lenne az eredményre jutáshoz (ezt Masoner tárgyalja az első és a második kísérletében, azonban ott nincs kiszűrve a lingvisztikai behatároltság a választásnál). Ez az ugyanabból a perspektívából felvett mozgáspárok közti különbség hasonló jelleggel áll fenn a külső nézetek esetében is.



17. Ábra: A DTW által kiszámított hasonlóság külső nézetű videók között

A cél ezen információk egybefoglalása és egy olyan módszer megtalálása volt amivel gyorsan meg lehet állapítani a tesztet kitöltők- és a számítógépes elemzés eredményei közötti korrelációt. Ezért bevezettük a DTWcompound nevezetű mérőszámot, ami egy a mozgáspárra jellemző mérőszám, és általa (legalábbis ezen esetekben) jól meghatározható melyik pár megkülönböztetése okozhat az embereknek nagyobb gondot.

A lentebbi diagramon látható a DTWcompound mindhárom párosítás esetében. Ez a mérőszám a következő egyenletből számítható:(egyenlet helye), Az egyenlet felépítését indokolja, hogy minél inkább hasonlít egymásra egy mozgás külső és belső perspektívából, annál könnyebb a párosítás. Szintén megkönnyíti a párosítást ha jobban különbözik a párjától. Így amennyiben egy mozgás két perspektívájából számolt DTW-hez hozzáadjuk ennek a DTW-nek és az egocentrikus videó és a párjához tartozó külső nézetes videó DTW értékének különbségét, akkor minél kisebb számot kapunk annál könnyebben meg lehet különböztetni és ezáltal párosítani a két videót.



18. Ábra: A párosítások DTWcompound értékei



19. Ábra: Masoner (2022) 3. kísérletének eredményei. A válaszadás pontossága (Accuracy) a különböző mozgáspárok függvényében

Végző konklúzióként látható, hogy a DTW compound és a mozgáspárok eltalálásának pontossága a Masoner kísérletben negatív korrelációt mutat egymással, vagyis minél jobban hasonlít egy mozgás két perspektívája és különbözik a párja másik perspektívájától annál pontosabban lehet beazonosítani a mozgásmintázatokat és ezáltal párosítani a különböző nézetű videókat.

4. Tárgyalás (Diszkusszió)

Kutatásunk során gondolkodtunk a deep learning, azaz a mesterséges intelligencia és neurális háló alapú módszerek alkalmazásán, hiszen a gépi látás területén jelenleg ezek a legkorszerűbb megoldások, ezek szolgáltatják a legjobb eredményeket. Ennek viszont két komoly akadálya is volt. Egyrészt, a dolgozatunk során egy már létező humán kísérletet próbáltunk számítógépes videóanalízissel igazolni és hasonlítani, így a vizsgált videók már előre adottak voltak. Ezek viszont nagyon kis adatmennyiséget jelentenek, egy egyszerűbb mély neurális háló tanításához is nagyságrendekkel több adatra lett volna szükségünk. Másrészt, a mély neurális hálók fekete doboz mivolta miatt nem lennénk képesek a mintázatok kiértékelésre. A háló által megtanult módszerek bizonyos mértékben visszafejthetők, azonban nehezen lennének számunkra értelmezhetők. Megvizsgálhatnánk, hogy milyen mozdulatokat tévesztene össze könnyen a háló, de arra, hogy a videóban szereplő vizuális információ mely részei alapján és hogyan hozza a döntéseit, nem kapnánk pontos választ.

A deep learning eszköztárának elvetésével viszont hátrányok is járnak. Se a mély neurális hálók, se emberek számára nem zavaróak az apró hibák és kis rendellenességek egy videóban, az agyunk többnyire túllendül ezeken a zavarásokon. Sajnos viszont erre egy szokásos algoritmus nem, vagy csak nagyon nehezen képes. Ezért a továbbiakban érdemes lenne, pár alapvető pontatlanságot kiszűrni az elemzett videókból, hogy tisztább mozgásmintázatot vizsgálhassunk.

Az algoritmusunk a pixelek mozgását érzékeli ezért nagyon fontos az egységes háttér, mert ugyanaz a mozgás két különböző háttér előtt teljesen más értékeket fog adni. (A függvények jellegre hasonlóak az értékek azonban jelentősen eltérhetnek.) Sok esetben a mozgás közben is változott a háttér. Ezt azért tudtuk elhanyagolni a kísérletekben, mert az összes videóban ugyanúgy megfigyelhető volt ez a zavaró jelenség. Azonban ez a pontosságot befolyásolja, ezért érdemes lenne a tesztet egységes háttér előtt felvett videókon is elvégezni. Ezentúl fontos lenne az, hogy az alkalmazott háttereken minél kevesebb csillogó felület legyen mert az a mérési eredményeinket szintén nagy mértékben torzíthatja. Szintén nagy fejlődési lehetőség lehetne, ha a belső és külső nézetből látott felvételeket egy mozdulatsoron vennék fel két különböző kamera állásból. Mivel csak egy kamera állt rendelkezésre, ezért a mozgásokat meg kellett ismételni, hogy minden szükséges nézetből rendelkezésre álljon a felvétel. A mostani apró különbségek a mozgások végbemenetele között zavarokat okoz a hasonlóság számításakor. Valamint sokat javíthatna elemzésünk pontosságán, ha több felvétel állna rendelkezésünkre.

A Masoner kutatásából átvett videókon a mozgásokat ugyanazon alany hajtja végre. Az adataink általánosíthatóságát javíthatnánk, ha különböző emberek mozgásáról is lennének felvételeink.

Az elemzett adatok javításán túl, a használt algoritmusokban és módszerekben is tudnánk változtatni, bővíteni. Pár zavaró jelenség kiszűrésével, és több videóanyag segítségével talán a korábban említett lag-korrelációs módszerrel is megfelelő minőségű eredményeket lehetne elérni.

Az algoritmusunkban az átlagolás megvalósításához a mozdulatokat azonos hosszúságúra nyújtottuk, vagyis idő szerint normalizáltuk az intenzitás-idő függvényeket. Mi történne viszont, ha a függőleges tengely irányában, tehát intenzitás szerint is végeznénk valamilyen fajta normalizációt? Két módszer hangzik ígéretesnek. Az első módszer ismertebb és széles körben használt a gyakorlatban: a standardizálás, vagyis az átlagintenzitás kivonása, majd osztás az intenzitásértékek szórásával. A második egyszerű osztás az átlagos intenzitásértékkel mellyel lényegében mellőzzük mennyire “mozgalmasan” végezte el a

mozdulatot az alany. További átgondolást igényel, hogy a kettő közül melyik lenne a megfelelő megoldás. Ha egy ilyen típusú normalizálást az idő szerinti normalizálással egyesítünk, azzal mind a mozgás sebességétől, erősségétől, kamerától való távolságától, és pár háttér-jellegű zavarástól eltekintünk. Ezzel a hátramaradt információ a mozgásnak vélhetően valamilyen mélyebb rendű mintázatát jelenti, amivel további érdekesnek ígérkező összehasonlítások végezhetők. Például segíthet több különböző alany esetén az egyéni mozgási sajátosságok elsímításában és egy általános mozgásleírás megtalálásában. Továbbá előnyös lehet egy metódus kidolgozása, amely a felvételeket minden nézőpontból származó adattal egyszerre számolva hasonlítja össze.

Az adatelemzések eddigi alapja a frame differenciálás volt, ami a vizsgált pixelek mozgásának csak az intenzitását írja le. Több videót felhasználva, és főként nem azonos mozgásirányú párokat vizsgálva érdemes lehet optic flow algoritmusokat használni, mert ezek segítségével nem csak a mozgás nagysága, hanem annak az iránya is megbecsülhető. Ezzel a bővebb információval más jellegű, összetettebb hasonlítások is lehetségessé válnak.

A videóelemzés eredményei a dolgozat alapján hozzájárulnak a humán észlelés alapkutatásához, hiszen statisztikailag alátámasztják Masoner kísérletének eredményeit. A módszer viszont a fentebb említett továbbfejlesztési lehetőségek fényében automatikus videóelemzésre még nem alkalmas.

5. Felhasznált irodalom

- [1] James J. Gibson, The ecological approach to visual perception. Boston: Houghton Mifflin, PP 332, 1979.
- [2] Johansson, G., Visual perception of biological motion and a model for its analysis. Perception & Psychophysics, 14(2), PP 201-211, 1973.
- [3] Masoner, H., Does Optic Flow Provide Information about Actions? Doctoral dissertation. 2022, <https://aquila.usm.edu/dissertations/1972>. (2022.11.07.)
- [4] Lishman, J. R., & Lee, D. N., The autonomy of visual kinaesthesia. Perception, 2(3), PP 287-294, 1973.
- [5] Rogers, B. J., Optic flow: Perceiving and acting in a 3-D world, i-Perception, 12(1), 1–25, 2021. <https://doi.org/10.1177/2041669520987257> (2022.11.07.)
- [6] Rogers, B. Young, K. Tootell, S., Optic flow and the maintenance of balance., Journal of Vision, 7(9):1023, 1023a, 2007. <http://journalofvision.org/7/9/1023/> (2022.11.07.)

- [7] Gibson, J. J., Optical motions and transformations as stimuli for visual perception. *Psychological Review*, 64, PP 228-295, 1957.
- [8] Sakoe, H., & Chiba, S., Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1), PP 43-49. 1978.
- [9] Vintsyuk, T. K., Speech discrimination by dynamic programming. *Cybernetics*, 4(1), 52-57., 1968.
- [10] McLeod, R. W., & Ross, H. E. Optic-Flow and Cognitive Factors in Time-to-Collision Estimates. *Perception*, 12(4), PP 417–423, 1983. <https://doi.org/10.1068/p120417> (2022.11.07.)
- [11] <https://www.pexels.com/video/young-guy-doing-break-dance-on-the-street-5362370/> (2022.11.07.)
- [12] Meijering, E. H. W., A chronology of interpolation: From ancient astronomy to modern signal and image processing, *Proceedings of the IEEE*, 90(3), PP 319–342, 2002.
- [13] https://www.usna.edu/Users/oceano/pguth/md_help/html/time0alq.htm (2022.11.07.)
- Megvalósítást segített:
- https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html (2022.11.07.)
- <https://dynamictimewarping.github.io/python> (2022.11.07.)