

Pannon Egyetem
Villamosmérnöki és Információs Rendszerek Tanszék
mérnök informatikus MSc

DIPLOMADOLGOZAT

Sztereó képek minőségi kérdései –
a diszparitás és a fixációs pontok kapcsolata

Kiss Péter József

Témavezető: Dr. Czúni László

2013

Nyilatkozatok

Alulírott Kiss Péter József diplomázó hallgató, kijelentem, hogy a diplomadolgozatot a Pannon Egyetem Villamosmérnöki és Információs Rendszerek Tanszékén készítettem mérnök informatikus MSc diploma (MSc in Computer Science and Engineering) megszerzése érdekében.

Kijelentem, hogy a diplomadolgozatban lévő érdemi rész saját munkám eredménye, az érdemi részen kívül csak a hivatkozott forrásokat (szakirodalom, eszközök, stb.) használtam fel. Tudomásul veszem, hogy a diplomadolgozatban foglalt eredményeket a Pannon Egyetem, valamint a feladatot kiíró szervezeti egység saját céljaira szabadon felhasználhatja.

Veszprém, 2013. május 06.

Alulírott Dr. Czúni László témavezető kijelentem, hogy a diplomadolgozatot Kiss Péter József a Pannon Egyetem Villamosmérnöki és Információs Rendszerek Tanszékén készítette mérnök informatikus MSc diploma (MSc in Computer Science and Engineering) megszerzése érdekében.

Kijelentem, hogy a diplomadolgozat védekre bocsátását engedélyezem.

Veszprém, 2013. május 06.

Köszönetnyilvánítás

Köszönetemet szeretném kifejezni témavezetőmnek, Dr. Czúni Lászlónak a segítségért és a hasznos tanácsokért, valamint Gál Mónikának és Lipovits Ágnesnek az adatok elemzésében való hasznos segítségükért!

A kutatási tevékenységet és az eredmények publikálását a Magyar Kutatási Alap támogatta (OTKA CNK 80368).

Tartalmi összefoglaló

A 3D technikák rohamos terjedésével előtérbe kerül a sztereó képek tömörítésének, illetve minőség értékelésének kérdése. A képminőség objektív méréséhez olyan módszerre van szükség, mely a lehető legjobban közelíti az emberi látórendszer tulajdonságait. A jelenleg két dimenziós képekhez használt objektív képminőség mérési jelzőszámok akkor alkalmazhatóak változtatás nélkül sztereó megjelenítés esetére is, ha a képmegjelenítési mód változása a fontos (fókuszba kerülő) képterületeken, illetve az érzékenység mértékén nem változtat.

Az ember számára fontos területek a képen számos különböző tényezőtől függenek például a nézési időtől, előképzettségtől vagy akár a hangulattól, szituációtól. E dolgozatban célom elsősorban a diszparitás fixációs pontok helyzetére gyakorolt hatásának vizsgálata. A diszparitás a két dimenzióban látott (monó) képhez képest, sztereó esetben többlet információként jelenik meg.

Dolgozatomban áttekintem a jelenleg alkalmazott képminőség értékelő jelzőszámokat, valamint bemutatom kutatásunkat, melynek célja a monó és sztereó képmegjelenítés esetén a tapasztalható fixációs pontok összehasonlítása. Célunk különbségek keresése a két képmegjelenítési mód között a szemlélő számára fontos területek tekintetében. A kutatás során két pszichovizuális kísérletet dolgoztunk ki.

A kísérletek során a résztvevőknek monó illetve sztereó módon természetes, illetve mesterségesen generált képeket jelenítettünk meg, mialatt a szemmozgást nyomon követve a nézett területekre következtettünk. Az eredményül kapott nézettségi mintákat (fixációs pontokat) statisztikai értelemben hasonlítottuk össze – célunk különbségek feltárása a monó illetve sztereó eredmények között.

Kulcsszavak: sztereó képek, képminőség mérés, fixációs pontok, vizuális kísérletek

Abstract

Due to the rapid spreading of 3D techniques have brought to the forefront stereo image compression and quality assessment issues. For the objective measurement of the image quality there is a need for a measurement method that approximate the best the capability of human vision. The objective metrics – that have been used recently – for measuring the quality of images for the two dimensional images can be applied without any modification for stereo images only in the case if some assumptions remain valid, namely the displaying mode of image has not been being altered on areas within the focus area (the saliency points), moreover the sensitivity has not been being modified.

The important areas for the human viewer (the saliency points) depend on many factors – the viewing time, prior training, and the situation. My goal in this paper was to study the role of disparity in the position of the fixations, which contains additional information in case of stereo display.

In my thesis, I have reviewed the current image quality evaluation metrics, and have presented our research, which aims to compare the mono and stereo display techniques. Our goal was to find differences at the location of fixation points in the saliency points.

In our research, we have developed two psycho-visual eye tracking experiments. In the experiments the participants were shown natural and artificially generated images in both display modes, and by tracking eye movement we could draw conclusions whether which areas has been being observed. We have compared the resulting fixation points statistically with the aim of identifying the differences between the mono and stereo results.

Keywords: stereo images, image quality assessment, fixation points, visual tests

Tartalomjegyzék

1. Bevezetés	10
2. Fogalmi áttekintés, elméleti háttér	12
2.1 Vizuális észlelés	12
2.1.1 A látórendszer és a látás folyamata	12
2.1.2 A látás néhány sajátossága	13
2.1.3 Térhatás észlelés	14
2.1.4 Térélmény sztereopszis alapján	16
2.2 Sztereó képek és megjelenítésük	17
2.2.1 Sztereó képek	17
2.2.2 Sztereó megjelenítés	17
2.2.3 A diszparitás szerepe a sztereó megjelenítésben	21
2.3 Képkódolás	22
2.3.1 Veszteségmentes kódolási eljárások	23
2.3.2 Veszteséges kép és videó kódolás	23
2.3.3 Sztereó képek kódolása	26
2.4 Képminőség és mérése	28
2.4.1 Képminőség értelmezése	28
2.4.2 Képminőség objektív és szubjektív mérése	29
2.4.3 Képminőséget jellemző jelzőszámok	30
2.5 Kiugró pontok („saliency”) és szerepük a képfeldolgozásban	36
3. Sztereó képekkel végzett vizsgálatok	39
3.1 Célkitűzés	39
3.2 Irodalmi áttekintés	41

3.3	Előzmények és korábbi tapasztalatok	42
3.3.1	Minőségromlás hatása a térérzetre	42
3.3.2	Következtetések	44
4.	Vizuális tesztek	45
4.1	Vizuális tesztek előkészítése	45
4.2	Felhasznált eszközök	46
4.2.1	Sztereo megjelenítés	47
4.2.2	Pupilla-követés	47
4.3	A teszt lefolyása	48
4.4	Rögzített adatok	49
4.5	Első vizuális teszt	50
4.5.1	Képi adatbázis	50
4.5.2	Résztvevők	51
4.5.3	Problémák, javítási lehetőségek	51
4.6	Második vizuális teszt	52
4.6.1	Képi adatbázis	52
4.6.2	Résztvevők	53
5.	Adatelemzés, eredmények	55
5.1	Szakkádok időbeli hosszának vizsgálata	56
5.2	Fixációs pontok időbeli hosszának vizsgálata	57
5.3	Mélységi határvonal távolság vizsgálat	58
5.4	Bejárt képterület vizsgálat	59
5.5	Diszparitás szinteken töltött idő eloszlásának vizsgálata	60
5.6	Egymáshoz legközelebbi fixációs pontok távolsága	61
5.7	Fix sugarú környezetbe eső fixációs pontok	63
6.	Szoftvertechnikai megvalósítás	66
6.1	Sztereo megjelenítő szoftver vizuális tesztekhez	66
6.1.1	Leonar3Do	67
6.1.2	OpenGL	67
6.1.3	OpenCV	67
6.2	Képgenerálás és adatfeldolgozás szoftvertechnológiai vonatkozásai	69

6.2.1	Mesterséges képek generálása	69
6.2.2	A kísérletek során rögzített adatok feldolgozásának támogatása	70
7.	Összefoglalás	72

1. Bevezetés

Egy kép már keletkezésekor az optikai rendszer, a CCD chip, illetve jelfeldolgozó, tömörítő eljárások számos hibalehetőségét hordozza magában, melyek a szemlélő számára zavaróak lehetnek. A kép egyes részei azonban különböző mértékben kerülnek a figyelem középpontjába, ezáltal a különböző képterületeken bekövetkezett hasonló mértékű torzulások az ember számára kialakult képminőség érzetet különböző mértékben befolyásolják. A képminőség értékelésre több minőség értékelő jelzőszám áll rendelkezésre monó képekre, mely automatikusan próbálja a képeken bekövetkezett torzulás ember által észlelt mértékét kifejezni.

A 3D technikák rohamos terjedésével azonban előtérbe kerül a sztereó képek tömörítésének és megjelenítésének, illetve minőség értékelésének kérdése. A képminőség értékelő jelzőszámok kialakításánál fontos kérdés az ember számára fontos képrészletek (sajátságos pontok) megállapítása, ahol a torzulásokra a szemlélő érzékenyebben reagálhat. A jelzőszámok sztereó képekre való alkalmazhatóságának ez által feltétele, hogy a sajátságos pontok helyzetén a képmegjelenítési mód változása ne nagymértékben változtasson.

Az ember számára fontos területek a képen sok különböző tényezőtől függnnek — a nézési időtől, előképzettségtől, szituációtól — pl. egészen másképp tekintünk meg egy képet egy keresési feladat esetén, mint ha csak egy véletlenszerű képkockát (például vágóképként) láthatunk. E dolgozatban célunk a kép jellemzőitől függő tényezők vizsgálata, különös tekintettel a diszparitás szerepére, mely a két dimenzióban látott (monó) képhez képest, sztereó esetben többlet információként jelenik meg.

Dolgozatomban áttekintem a jelenleg alkalmazott képminőség értékelő jelzőszámokat, valamint bemutatom kutatásunkat, melynek célja a monó és sztereó képmegjelenítés esetén a tapasztalható fixációs pontok összehasonlítása, különbsé-

gek keresése a két képmegjelenítési mód között a szemlélő számára fontos területek tekintetében. A kutatás során két pszichovizuális kísérletet dolgoztunk ki, melyek során a résztvevőknek monó illetve sztereó módon természetes, illetve mesterségesen generált képeket jelenítettünk meg, mialatt a szemmozgást nyomon követve a nézett területekre következtettünk. Az eredményül kapott nézettségi mintákat (fixációs pontokat) statisztikai értelemben hasonlítottuk össze — célunk különbségek feltárása a monó illetve sztereó eredmények között.

2. Fogalmi áttekintés, elméleti háttér

2.1. Vizuális észlelés

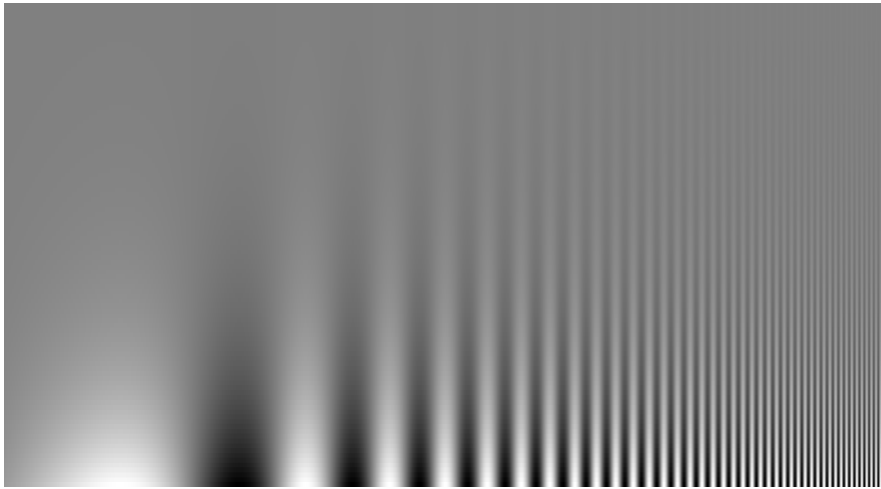
2.1.1. A látórendszer és a látás folyamata

A látás folyamata alapvetően két szakaszra bontható – az érzékelésre és az észlelésre. Az emberi fényérzékelés a szem feladata. A szem hátsó belső oldalán az érhálózat és idegsejtek mögött találhatóak a retina kétféle fényérzékeny sejtjei. A csapok illetve a pálcikák felelősek a fényingerek ingerületekké alakulásáért. A csapok kisebb érzékenyséűek, a pupillával szemben a sárgafolt közelében helyezkednek el és három korlátozott hullámhossz-tartományban érzékenyek (három fajta csap három különböző tartományban), ezáltal nagy fényerősség esetén képesek vagyunk a színek megkülönböztetésére – a színészlelet a különböző fajta csapok különbségi információiból az agyban alakul ki. A nagyobb érzékenyséű pálcikák, amelyek sokkal nagyobb számban gyakorlatilag – a vakfolt és a sárgafolt kivételével – a retina teljes területén megtalálhatóak, a kis fényerősségekben történő éjszakai látásért felelősek.

A szem izmai az emberi test egyik leggyorsabb mozgására képes részei, melyekkel tekintetünket egyik pontról a másikra tudjuk irányítani. Ezen gyors mozgásokat szakkádoknak nevezzük, melyek során az objektumok vetülete a perifériáról a sárgafolt közelébe kerül a retinán. A szem egy bizonyos ponton történő megnyugvását fixációnak, ezen helyeket a képen fixációs pontnak nevezzük. A látás során az ugrások (szakkádok) és megnyugvások (fixációk) felváltva követik egymást. A szakkádok során információ felvétel nem történik, ezt a szakkádikus elnyomás

jelenségének nevezzük, vagyis a fontos képrészletek megállapításához a fixációs pontok vizsgálata elégséges. (A szakkádikus elnyomás jelenségéről egy érdekes kísérlet segítségével is meggyőződhetünk – ha a tükör előtt állva fejmozdítás nélkül tekintetünket egyik szemgolyónkról a másikra helyezzük át, mozgást sosem fogunk látni.)

A látható optikai sugárzás tehát a szem speciális sejtjei segítségével alakul érzetté, ingerületté, azonban a látás észlelet kialakulásáért elsősorban az agy a felelős.



2.1. ábra. A kontraszt érzékenységi függvény szemléltetése

2.1.2. A látás néhány sajátossága

Az emberi érzékelésre és észlelet kialakulására általánosan elfogadott törvény Weber törvénye, mely szerint az észlelhető intenzitás különbség az észlelt jel valós intenzitásával egyenesen arányos, vagyis minél nagyobb az észlelt jel intenzitása, az éppen észlelhető különbség, vagyis az észlelhető legkisebb változás mértéke annál nagyobb lesz.

A szem fizikai felépítése, neurális működése, illetve a vizuális ingerületek agy által végzett feldolgozása emellett számos sajátosságot hordoz magán, mely az észleletet befolyásolja, e sajátosságokat elsősorban a művészetben az ókortól kezdve kihasználták (pl. görög templomok torzított perspektívája), s talán leginkább optikai csalódások néven váltak ismertté. Ilyen sajátosság, hogy az úgynevezett kont-

raszt érzékenységi függvény értéke nem konstans a frekvencia függvényében, vagyis a képen éppen érzékelhető kontraszt változás értéke a frekvencia függvényében változik. Ezt magunk is érzékelhetjük a 2.1. ábra segítségével ahol a függőleges tengelyen a kontrasztot csökkenettük a függőleges tengely mentén, és ábrázoltuk a frekvencia függvényében – látható, hogy az érzékenységi küszöb (a homogénnek ítélt terület határvonala) nem vízszintes. A kontraszt érzékenységi függvény az adott képhez és környezethez való alkalmazkodás, adaptáció során időben kis mértékben változhat.



2.2. ábra. Szimultán kontraszt

Hasonló sajátosság, hogy egy adott terület fényességének értékét az agy nem önmagában vizsgálja, a fényesség észlelete a környezet függvényében alakul ki, ezt szimultán kontraszt jelenségnek nevezzük. A 2.2 ábrán látható homogén intenzitású, fénysűrűségű területet a háttér változása miatt különböző intenzitásúnak észleljük.

Hasonló hatások színekkel és formákkal is kimutathatók, bizonyos képekkel akár a mozgás illúziója is kelthető. Az emberi látás komplex folyamat, az agy az észlelet során magas szintű feldolgozásokat is végez, így sok sajátossággal rendelkezik, melyek részletes bemutatása túlmutat e dolgozat keretein.

2.1.3. Térhatás észlelés

A térlátást nem lehet egyszerű fiziológiai érzékelésként felfogni, a távolság észlelés nem közvetlenül receptorok segítségével történik, hanem különböző érzetek, részinformációk hatására az agyban alakul ki. A térélményért részben a szemlencse fókuszálása és mélységélessége a felelős – amit éppen nézünk élesen, a közelebbi (illetve távolabbi) objektumokat homályosabban látjuk. Két dimenziós képeken is



2.3. ábra. Perspektíva

tapasztalhatunk némi térélményt, meg tudjuk különböztetni az előteret a háttértől, képesek vagyunk a távolság- és méretarányok megállapítására, emellett egy szemmel is képesek vagyunk a távolság korlátozott meghatározására. Egy 2D képpel is kelthető a távolság illúziója, például perspektivikus hatások következtében (2.3. ábra) [1].

Az emberi látásról tapasztalatból tudhatjuk emellett, hogy a két szem egymástól különböző képeket képez le. Ennek belátására egy egyszerű alapkísérlet végezhető el: felváltva letakarjuk a jobb és bal szemünket, közben nézzük az orrunk előtt kis távolságban tartott ujjunkat. Az ujjunk elmozdulni látszik [39]. Azt is tapasztalhatjuk, hogy ha tartósan csak egyik szemünkkel nézünk (a másikat letakarjuk), a tájékozódásban csak kismértékű zavarokat tapasztalunk, de bizonyos térbeli műveletek már érezhetően nehezebben mennek (például tűbefűzés). A térérzet tehát részinformációk halmazából az agyban jön létre. A térélményhez a távolság észlelése szükséges, mind a szemlélőtől való abszolút, mind a képen látható két objektum közötti relatív távolságot ideértve. Az emberi távolság-észlelésnél a távolsági jelzőmozzanatoknak van jelentős szerepe (szemmozgásos-, látási-, monokuláris-, binokuláris jelzőmozzanatok) [29, 12]. A sztereó megjelení-

tésben leginkább kihasznált távolság-észlelési folyamatot a binokuláris jelzőmozzanatok valósítják meg. A két szem ugyanis – elhelyezkedéséből adódóan – a tárgyakat kissé eltérő pozícióban látja (az optikai tengelyhez képest), ezt a jelenséget nevezzük retinális diszparitásnak. Az agy a két kép eltéréséből tud következtetni a távolsági adatokra – a diszparitás szerepét a térérzet kialakulásában bővebben a 2.2.3. pontban mutatom be.

2.1.4. Térélmény sztereopszis alapján

Sztereopszis alatt azt a vizuális észlelési folyamatot értjük, mely a két szem által alkotott két különböző vetületű kép alapján a mélység érzetét nyújtja. (Az elnevezés a *stereo* „valódiság” és az *opsis* „látvány vagy látomás” szóösszetételből származik [14, 41]). Lényeges kérdés, hogy létezik-e önállóan a sztereopszis szolgáló binokuláris térlátás – vagyis kizárólag a képek vetület különbségéből is jó minőségű térhatás észlelet alakul ki. Ennek bizonyítására szolgáltak az úgynevezett randompont-sztereogrammal kapcsolatos kísérletek. A randompont-sztereógrammot elsőként Julesz Béla fejlesztette ki. A randompont-sztereogramokon egy véletlen ponthalmazt tartalmazó képből, a kép egy kijelölt darabját (pl. egy geometriai objektum területét) kis mértékben elmozdítva állíthatjuk elő a másik szem képét [16]. Külön-külön tekintve a képeken a kijelölt objektum nem jelenik meg, a két képet egyszerre megjelenítve (sztereoszkóp segítségével) a kijelölt forma kiemelkedik a háttérből, attól különböző távolságban észlelhető. A randompont-sztereogrammok bizonyították, hogy önállóan a binokuláris térlátás is létezik, nincs szükség a térélményhez a kép tartalomtól függő monokuláris jelzőmozzanatokra – a sztereopszis önállóan is alkalmazható 3D megjelenítésre. (A kísérleteinkben résztvevő személyeknél randompont-sztereogramokat alkalmaztuk a helyes sztereó látás ellenőrzésére).

A sztereopszis alapú térélményhez alapvetően a binokuláris jelzőmozzanatokat használjuk ki, vagyis a két szem által látott különböző képek alapján alakul ki a térhatás észlelet. A megfelelő térélmény kialakulásához tehát azt kell biztosítani, hogy mindkét szem csak a neki szánt képet lássa, vagyis a két szemnek szánt képeket valamilyen műszaki megoldással el kell választani egymástól.

2.2. Sztereó képek és megjelenítésük

2.2.1. Sztereó képek

A sztereó képpárok, mivel ugyanazt a témát tartalmazzák két különböző vetületből általában (szélsőséges esetektől eltekintve) nagy mennyiségű redundáns (ismétlődő) információt tartalmaznak – a két képen nagy részben azonos információ van jelen, csak esetlegesen eltérő helyzetben, illetve eltérő fizikai helyen. E redundancia két szempontból is fontos lehet vizsgálódásunk folyamán – a 3D adattárolás illetve adattovábbítás esetén azonos felbontás biztosítása esetén a szükséges adatmennyiség megkétszereződik, így hatékony adattömörítési eljárásokra van szükség. Ebből a célból az elmúlt néhány évben létrehozták az MPEG ún. több nézőpontú verzióját [32]. Ugyanakkor felmerülhet az a kérdés is, hogy a térérzet helyes kialakulásához szükséges-e mindkét kép teljes információtartalma, vagy az agy képes bizonyos hiányosságokat korrigálni.

2.2.2. Sztereó megjelenítés

A sztereó képeket általában normál (sík) kijelzőn jelenítik meg. A megjelenítés feltétele, hogy mindkét szem csak a maga képét láthassa. Bár ismert a Gábor Dénes által felfedezett holográfia [10] alkalmazása a térhatású képek megjelenítésében, annak speciális technikai feltételrendszere miatt ritkán alkalmazzák – a sztereó képjelenítésben a két (jobb és bal szemnek szánt) képet megjelenítő eljárások terjedtek el. A holografikus megjelenítéssel kutatásunk során nem foglalkoztunk.

2.2.2.1. Sztereográf

A sztereografikus képnek a szem átlagos távolságára (63 mm) elhelyezett két (egymástól ugyanilyen távolsággal fényképezett) képet nevezünk. Itt a térélményt a szemlélőnek kellett „kiprovokálnia” azzal, hogy szeméhez közel tartva a képet megpróbál a messzeségbe tekinteni. A távolra fókuszálással a két szem lényegében a neki szánt képpel néz szembe, így az agyban a térhatás kialakulhat.

2.2.2.2. Sztereoszkóp

A sztereoszkóp olyan speciális (többnyire tükrökkel megvalósított) eszköz, ami biztosítja, hogy a két szem mindegyike csak a neki szánt képet láthatja – a sztereográfnál kényelmesebb megjelenítést biztosít, illetve segítségével élesebb képeket láthatunk. Hosszú ideig elterjedt, vizuális kísérleteknél ma is előforduló sztereó megjelenítési technika.

2.2.2.3. Színszűrős szemüveges technikák

A mechanikai eszközökhöz képest jelentős előrelépést jelentett a színszűrős (úgynevezett *anaglif*) technikák megjelenése. A színszűrős technika lényege egy olyan szemüveg, amelynek a jobb és baloldali lencséje eltérő karakterisztikájú színszűrővel van ellátva, ezáltal lehetővé válik a két szemnek szánt különböző látvány egy fizikai képen történő megjelenítése. A jelenleg is leginkább használatos színpárok a vörös - cián, és a magenta - zöld, de alkalmazzák a kék - borostyán változatot is. Az anaglif technikánál egy közös megjelenítőn a két képet színátalakítással jelenítik meg úgy, hogy a egyik képet csak az egyik szem előtt található színszűrő, a másikat csak a másik engedje át. A megoldás nagy előnye, hogy olcsó és nyomtatásban is használható – az akár papírszemüvegekkel is alkalmazható technika szélesebb körben tette elérhetővé a sztereó megjelenítést. Hátránya viszont – azon túl, hogy szemüveg kell hozzá – a számottevő színtorzítás (nem adható vissza maradéktalanul az összes szín és árnyalat, illetve a féhéregyensúlyt is eltolja). Az agy számára is meglehetősen munkaigényes az eltérő színtónusú képek feldolgozása, így a tapasztalatok szerint a szemüveg viselése hosszú távon fárasztó [30]. A színszűrők nem tökéletes vágási karakterisztikája miatt az alkalmazási gyakorlatban szellemkép-hatás is megjelenik. Anaglif képek készítésére és a hibák kiküszöbölésére több megoldás létezik (pl. színcsatorna információk keverése – a vörös kép ne csak a bal kép vörös csatornájából készüljön), melyek általában egyik hibán segítenek, azonban más tulajdonságokat rontanak.

2.2.2.4. Polarizációs technikák

A polarizációs technikák a két kép szétválasztását a fény polarizációs irányának változtatásával oldják meg, a két szemhez eltérő, egymásra merőleges polarizációs

írányt választva, és ennek megfelelő polarizált fénnel való megjelenítést, illetve szűrőt alkalmazva elvben színhelyes, bár néha kicsit sötétebb képet kapunk.

Probléma, hogy a fix helyzetű megjelenítőhöz képest a fejet elmozgatva (elforgatva) a polarizációs iránytól elmozdulunk, így a másik szem előtt lévő szűrő is átenged és a kép szellemképpé válik. Ezt a hátrányt egy speciális polarizációs technika, az úgynevezett körkörös polarizáció alkalmazásával korrigálni tudják [5]. A módszer alkalmazása esetén a fény polarizációs síkját folyamatosan (az idő függvényében, a két szemnek szánt képeknél egymástól ellentétes irányban) elforgatjuk, ezáltal két térben polarizált hullámot kapunk, melyekből az adott szemnek szánt képet egy negyedhullám-lemez (quarter wave plate) és egy lineáris polárszűrő együttesével tudjuk kiszűrni. A módszer alkalmazásával a szemüveg bármely helyzetében elkülöníthetők a két megjelenítőből származó fénysugarak.

Polarizációs technikával dolgozik például az Imax3D mozi kép megjelenítése, melynél a felhasználási környezet kialakítását igencsak drágává teszi, annak biztosítása, hogy mozivásznonról a diffúz visszaverődés polarizációs-irány tartó legyen – vagyis a két független megjelenítő által kibocsájtott képek polárszűrő segítségével a vászonról való visszaverődés után is szétválaszthatóak legyenek.

2.2.2.5. Aktív szemüveges technikák

A legújabb generációja a szemüvegeknek nem egy adott frekvenciasávot, vagy polaritást szűr, hanem nyit/zár alapon szeparál. Itt egy — jelenleg folyadékkristályos — szemüveg panel a vezérléstől függően átlátszik, vagy teljesen sötét. A megjelenítő ekkor a két szemnek szánt képet felváltva jeleníti meg, s ezzel szinkronban nyit/zár a szemüveg. A felhasználó tehát felváltva hol a jobb, hol a bal szem képét láthatja (jobb vagy bal szemével). Természetesen a korábban használt megjelenítési képfrekvencia (50Hz) nem elegendő, mert mindig csak az egyik szem kap képet, így az egy szemnek jutó megjelenítési frekvencia gyakorlatilag feleződik, 25Hz megjelenítési frekvenciánál pedig az emberi látás számára erős villogást érzékelhető. Az aktív szemüveges rendszereknél a megjelenítőtől minimum 100 Hz képfrissítési frekvencia elvárható (a gyakorlatban általában 120 Hz az elvárás, mely biztosítja mindkét szem számára a villogásmentes, hosszabb távon is kényelmes megjelenítést). Aktív technikánál a megjelenítővel szemben további fontos követelmény, hogy a képváltást követően elegendően kicsi legyen az utánvilágí-

tási idő (válaszidő). Az aktív szemüveges technikák tehát speciális megjelenítőt, illetve a gyors képfrissítés miatt erős grafikai hardvert (grafikus kártyát) igényelnek, mindez használatukat egyelőre meglehetősen költségessé teszi.

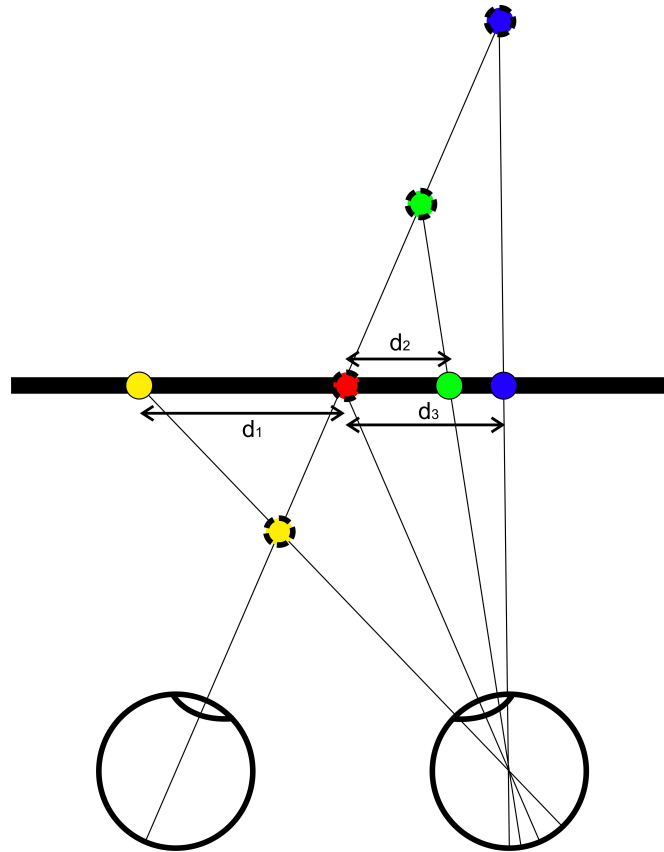
Külön említést érdemel egy hazai fejlesztésű aktív szemüveggel ellátott eszköz, a Leonar3do [19], mely valójában egy virtuális valóság rendszer, ami nem csak megjelenítést, hanem a térbeli munkamenetet is támogatja. A speciális (aktív folyadékkristályos) szemüveg és a beavatkozó mutató („madár”) mozgását infravörös érzékelők figyelik, az eszköz ezekkel biztosítja a 3D térben való munka érzését. A Leonar3Do nagy újtása, hogy figyeli a megtekintő helyzetét (a szemüveg, ezáltal a szemek helyzetét a monitor előtti térben), így a szemlélő helyzetének változására a megjelenített képet változtatni tudja, melynek segítségével teljesebb virtuális valóság hatás érhető el, grafikai objektumok esetén akár azok mögé is tekinthetünk.

2.2.2.6. Autosztereoszkópia

Napjainkban egyre nagyobb igény van a felhasználó oldali segédeszköz nélküli (auto-sztereoszkopikus) 3D megjelenítésekre. Részben eltérő módszerek alakulnak ki a sztereoszkopikus (két képpel dolgozó), és a multi-view, több nézőpontú (alapvetően szélesebb nézéstartományú) megközelítések között. Megjelentek a néző szemét, vagy arcát figyelő, övető megoldások, amelyek alkalmazkodóképes, még valósághűbb térérzetet biztosítanak [8]. Egy-egy játék, vagy munkavégzés esetén ugyan (néha kényszerből) elfogadott a speciális szemüveg, de már a szórakozásnál szükség lenne speciális eszköz nélküli 3D képek biztosítására. Az autosztereoszkópia a speciális segédeszköz nélkül biztosított térélményt jelenti, nevezik „szemüvegmentes” 3D technikának is, melynél két fő technikai megoldás terjedt el.

A rácsmaszkos megoldásnál a szembe jutó fény egy a megjelenítő elé helyezett ráccsal maszkolódik úgy, hogy mindegyik szem csak a neki szóló pixeleket látja. Ugyanezt éri el az ún. lenticuláris képernyő, ahol a megjelenítő elé helyezett lencséalakja biztosítja a képek szétválasztását.

„Multi-view” rendszerek esetén, ahol kettőnél több képet vesznek fel, és értelemszerűen több képet jelenítenek meg a térélmény kiterjesztéséhez (a nézőpont kismértékű változtatására érzékeny megjelenítési formát hozva létre — elmozdulva kicsit más képet láthatunk — mellyel a térhatás nagymértékben javítható) [8].



2.4. ábra. A diszparitás hatása térérzetre. A bal szemnek szánt kép a piros színű helyre rögzítve, a jobb szem képe különböző diszparitás értékekkel eltolva.

2.2.3. A diszparitás szerepe a sztereó megjelenítésben

Sztereopszison alapuló sztereó képmegjelenítés esetén a távolság észleletet elsősorban a két képen az objektumok eltérő helyzete okozza. A két képen az objektumok, illetve azok egyes alapegységeinek (digitális képek esetében pixeleinek) ezt az eltolását, távolságát értjük diszparitás alatt. A diszparitás nagyságától és irányától (előjelétől) függően az objektumok a képmegjelenítő síkjában (0 diszparitás), illetve az előtt, vagy mögött észlelhetők melyeket a 2.4. ábrán szeléltetnek. A bal szem képét minden esetben a piros színnel jelölt helyre rögzítettük. Amennyiben a jobb szemnek szánt kép is ugyanezen a helyen kerül megjelenítésre az objektumot a képernyő síkjában látjuk megjelenni. Ha a jobb szemnek szánt képen az objektumot különböző mértékben eltolva (különböző diszparitásokkal) jelenítjük meg, az objektumot a képmegjelenítő előtt vagy mögött észleljük.

Látható, hogy a különböző diszparitásokkal megjelenített objektumok képe a retinán is különböző helyeken jelennek meg. A retinára vetülő képek szemtengetyeltől való távolságkülönbségét retinális diszparitásnak nevezzük – a diszparitás fogalma tehát természetes látásban is megjelenik.

2.3. Képkódolás

A képeket nem csak felvenni és megjeleníteni, hatékonyan tárolni és továbbítani is kell. A digitális álló és mozgóképek a számítástechnikában előforduló egyéb sűrűn használt adatokhoz képest nagyságrendileg nagyobb tárhelyigénnyel rendelkeznek. A monokromatikus (egyszínű pl. szürkeskálás) képeket általában a képtérben, pixelenként (képpontonként) reprezentáljuk, azaz pixelenként egy-egy fényesség értéket tárolunk el. Az általános felhasználásra szánt képekre általában pixelenként 8 bit áll rendelkezésre a fényesség értékek kifejezésére, mely 256 különböző érték tárolását teszi lehetővé. Speciális alkalmazásokban, pl. orvosi képfelvételek esetében nagyobb bitmélységet alkalmaznak, ezzel növelve a színskála felbontását. A színes képek tárolásánál az additív színkeverés elve alkalmazható, vagyis a vörös, kék és zöld fényesség értékeket külön-külön kell eltárolni. Egy alacsony felbontású, körülbelül a TV felvételnek megfelelő minőségű színes kép 512×512 -es felbontásához pixelenként három szín tárolása esetén $512 \times 512 \times 8 \times 3 = 6.3 \times 10^6$ bite lesz szükség [27], mely kb. 0.75 MB tárterületet jelent. Ha egy ilyen képekből álló szokványos videó folyam (25 kép másodpercenként) minden egyes képét ilyen módon tárolnánk el, egy hagyományos DVD lemezre mindösszesen kb. 4 percnyi videót tudnánk rögzíteni ($0.75\text{MB} \times 25\text{f/s} \times 60\text{s} \times 4 = 4500\text{MB}$).

A nagyfelbontású kamerák és megjelenítők terjedésével a képek és videók adatmennyisége megsokszorozódik, a HDTV által használt 1920×1080 -as felbontás képenkénti adatmennyisége kb. 8-szorosa az előbb bemutatott példának. A sztereó és multiview technikák pedig egy-egy időpillanathoz egy helyett nézetenként egy, azaz kettő vagy több kép tárolását teszik szükségessé. Érezhető, hogy a felhasználói élmény, a felbontás és minőség növeléséhez szükséges a képek adat-reprezentációjának hatékony tömörítése.

2.3.1. Veszteségmentes kódolási eljárások

A veszteségmentes tömörítési eljárások a jelfeldolgozásban, kommunikációban és adattárolásban széles körben ismert és alkalmazott eljárások [38], melyek egy adott bitmintát, bitsorozatot tömörebben, azaz kevesebb bit segítségével képesek eltárolni. Ezek az eljárások nem veszik figyelembe a tömörítendő adat jellegét, pusztán 0-k és 1-esek sorozatát alakítják kisebb tárhelyigényű formába úgy, hogy az adatok a későbbiekben teljes mértékben helyreállíthatóak legyenek, vagyis a tömörítés bemenete, és a kitömörítés kimenete megegyezik. A tömörítéshez az adatok bináris mintáiban található redundanciát aknázzák ki. Veszteségmentes kódolás például a futam-hossz kódolás, mely során a tömörítő eljárás a 0-k és 1-esek értékeinek letárolása helyett, egy érték - számláló párt tárol. Hosszabb 0 vagy 1 sorozatok esetén jelentős helymegtakarítást érhetünk el. Hasonló tömörítő eljárás a Huffman kódolás, mely a sűrűn előforduló kódszavakhoz rövidebb, a ritkábban előfordulóakhoz hosszabb kódszavakat társít, így tárterület takarítható meg. A veszteségmentes kódolók azonban nem tudnak az előbb említett fejlődési tendenciának megfelelő tömörítési hatékonyságot biztosítani.

2.3.2. Veszteséges kép és videó kódolás

A veszteséges tömörítési eljárások lényege, hogy a tömörítő bemenete és a kitömörítés kimenete nem egyezik meg – kis mennyiségű, lehetőleg alig észlelhető információ veszteség mellett azonban nagy mennyiségű tárhelyet takaríthatunk meg. A veszteséges tömörítési eljárások általában figyelembe veszik a tömörítendő adatok eredeti megjelenési formáját (képről, hangról, videóról van-e szó) és az emberi érzékelést, észlelést is – esetünkben az emberi látás fiziológiai sajátosságait.

2.3.2.1. A JPEG képtömörítő szabvány

A JPEG-kódolás állóképek, fényképek tömörítésére szolgál. Alapja, hogy az emberi látás a világosságra érzékenyebb, mint a színekre, valamint a képen szereplő frekvencia komponensek sem azonos módon észlelhetőek. A JPEG szabvány számos működési módot és opciót tartalmaz, e dolgozat keretében csak az általánosan használt veszteséges soros módot ismertetjük.

A JPEG kép kódolásának első lépéseként a szokásos csatornánkénti RGB reprezentációból áttérünk a YCrCb színtérbe, mely egy világosság csatornát (Y) és két szín-különbségi csatornát tartalmaz. A világosság csatorna információinak változatlanul tartása mellett a színcsatornák esetében kettő, vagy négy pixeles blokkok értékét átlagoljuk, ezáltal a színcsatornák felbontása felére, vagy negyedére csökken (opciótól függően). Ezt a veszteséges lépést az ember az érzékenységi különbségek miatt alig észleli. Ezután egy blokk-előkészítési lépés következik, melyben a képet 8×8 -as blokkokra osztjuk fel. A blokkokon DCT-t, azaz diszkrét koszinusz transzformációt hajtunk végre. A transzformáció eredménye egy egyenérték (átlagérték), illetve a képen szereplő különböző frekvencia komponensek együtthatói. A DCT elméletben veszteségmentes, azonban a lebegőpontos számábrázolásból adódóan kismértékű torzulás előfordulhat. A DCT együtthatókat tartalmazó mátrixot ezután egy, a szabványban meghatározott kvantálási mátrix értékeivel elosztjuk, majd pedig egészre kerekítjük, mely lépésekkel gyakorlatilag kitöröljük a nagyobb térbeli frekvenciás komponenseket. Az egyenérték ezután az előző blokk azonos értékétől való eltéréssel helyettesíthető, mely várhatóan a lassan változó értékek miatt kisebb szám lesz az eredeti értéknél. A DCT együtthatók kvantált értékeit ezután egy cikk cakk bejárás szerint kiolvassuk és a 2.3.2. pontban említett futam hossz, illetve Huffman kódolást alkalmazzuk.

A JPEG kép dekódolásához az algoritmus fordítottját kell lefuttatni. A kódolás és dekódolás idő szükséglete nagyjából szimmetrikus. A JPEG kódolás eredményül kapott kép mérete a kódolási paraméterektől és a működési módtól függően akár az eredeti kép 20-ad részére is csökkenhet (legjobb minőség alkalmazása esetén az eredeti 24 bit/pixel helyett átlagosan 8.25 bit/pixel tárterület szükséges). Nagymértékű JPEG tömörítés esetén azonban a blokkonkénti feldolgozás miatt a lassú átmenetes területeken a blokk-határoknak megfelelő új élek jelenhetnek meg, illetve „blokkosodás” figyelhető meg.

A JPEG szabványnak több kiegészítése és továbbfejlesztése létezik, melyekkel hatékonyabb tömörítést, vagy jobb minőség érzetet lehet elérni, ilyen például a szabványban szereplő diszkrét koszinusz transzformáció helyettesítése az emberi látás sajátosságaihoz jobban illeszkedő Wavelet-traszformációval.[6]

2.3.2.2. Az MPEG videótömörítő szabvány

Az MPEG szabvány célja mozgóképek hatékony tömörítése. A videó formátumok, mozgóképek gyakorlatilag állóképek sorozatát tartalmazzák, amelyeket elég gyorsan egymás után vetítve az ember folyamatos mozgásként fog észlelni. Felmerülhet a kérdés, hogy ha a JPEG bevált az állóképek tömörítésénél, akkor miért nem elég a képsorozat minden képét JPEG kódolással tömöríteni, majd az így kapott sorozatot újból összefűzni? Természetesen ezzel a módszerrel is már meglehetősen nagy tárhely-igény csökkenést lehet elérni, azonban a szokásos videó képsorozatok rengeteg időbeli redundanciát tartalmaznak – a képtartalom nagy része két szomszédos képkockán általában változatlan, vagy kis mértékben mozdult el. Az MPEG ezt az időbeli redundanciát használja ki a videók tömörítésére. Az MPEG által tömörített videó folyamában négy fajta képet különböztethetünk meg.

- I (Intracoded) képek: önállóan JPEG kódolt képek (csak saját információ figyelembe vétele)
- P (Predictive) képek: blokkonként eltérés számítása az előző képtől (az előző kép adatainak figyelembe vétele)
- B (Bidirectional) képek: blokkonként eltérés az előző és a következő képtől (előző és következő kép adatainak figyelembe vétele)
- D (DC-kódolt) képek: blokkátlagokat tartalmazó képek – kiegészítő kiegészítő kiegészítő képek, a videóban való keresés megkönnyítésére (gyorsan megjeleníthetőek), melyek az MPEG-2 szabványtól kezdve nincsenek támogatva

Fontos megjegyezni, hogy az MPEG általában 16×16 -os makroblokkokat ért blokkméreten. A kódolás során a makroblokkok helyzetét vizsgáljuk a két szomszédos képkockán, a szükséges eltolás mértékét kiszámítva tlaajdonképpen a mozgásvektort tároljuk. Az előző képkockából a mozgásvektorokkal kiszámolt kép, illetve a kódolandó képkocka különbségét külön kódolva hozzáadjuk a képtartalomhoz. A különbségi képek tárhely igénye természetesen jóval kisebb az önállóan kódolt I képekhez képest. Hatékonynak tűnhet tehát az első kép után a többi különbségi elven kódolni, azonban ez a kódolást meglehetősen sérülékennyé tenné, illetve a videóban való mozgást (előre, hátra tekerés) erősen akadályozná. Az I

B P képek arányát a felhasználás célja, a csatorna és a tárhely biztonságossága határozhatják meg. Amennyiben minden képhez egyedi hozzáférés szükséges (pl. videó vágásnál), használnak csupa I képből álló, azaz önállóan képenként JPEG kódolt videó folyamatot is.

2.3.3. Sztereó képek kódolása

A 3D-vel kapcsolatos adattárolásnál tekintettel kell lennünk arra, hogy a sztereoszkopikus megjelenítésből lényegében a hagyományos videojel-továbbítás adattömegének a megkétszereződése adódik. A „multi-view”, azaz több nézőpontos rendszereknél az adatok mennyisége arányosan nő a megjelenített különböző képek (a lehetséges különböző nézőpontok) számával. A szükséges adatok mennyiségéből, illetve a képek méretéből adódóik, hogy a 3D jelfolyamot támogató formátumok esetében az egyik kulcskérdés a hatékony tömörítés. Itt érdemes figyelembe venni, hogy jelentős mértékű redundáns információval rendelkeznek a sztereó, valamint a multi-view képek, (lévén eltolásokkal nagyrészt ugyan azt a témát ábrázolják). A sztereó képek hatékony tömörítéséhez tehát jól kihasználható a redundancia. Sztereó videófolyamnál az említett redundancia a bal és jobb szem képe között térben, mind időben (az előző és következő, illetve az adott képkocka között) megjelenik.

A szabványosítást illetően ehhez kapcsolódva [40] azt állapítja meg, hogy egyelőre vezető helyzetben lévőknek a HDMI 1.4 specifikációt tekinthetjük, amely a szabványba emelte a tömörítetlen 3D formátumoknak a natív támogatását. Ezen a területen kiemelt fontosságú, hogy megőrződjön a meglévő technológiákkal való kompatibilitás, azaz a hagyományos megjelenítők alkalmazhatók maradjanak. Ehhez a váltott soros jelfolyam alkalmazása az egyik alaps megoldás, mivel ez alkalmazható a hagyományos megjelenítőkön is. E megoldásban a kép páratlan és páros sorai szolgálnak a bal és jobb szem képalkotásához forrásként (lehetséges más elrendezés is, például X alakban vagy oszloponként váltás). Egy másik lehetőség az úgynevezett időmultiplexált változat, ahol a bal és jobb szem számára szolgáló képek a jelfolyamban egymást követik, természetesen megnövelt (másodpercenkénti) képszámmal. Előnye e megoldási változatoknak, hogy teljes mértékben a ma elterjedt infrastruktúrára (adathordozókra, jeltovábbítási technológiára, formátumokra) épülnek, azokkal teljesen kompatibilisek, azaz a jelenleg hozzáférhető

készülékek (minimális változtatásokkal) sztereoszkopikussá tehetők. A technológia korlátai miatt viszont hátránya, hogy nagy mértékben korlátozza (leszűkíti) a 3D technikában rejlő lehetőségek kihasználhatóságát.

Jelenleg a legelterjedtebb bizonyos értelemben „legjobb” szabványnak a videó-kódolásban ma az ISO/IEC 23002-3 (MPEG-C Part 3), illetve a H.264/MPEG-4 AVC szabvány tekinthető. Az előbbi a mélységinformáció különálló reprezentációjára, utóbbi pedig a váltott soros és időmultiplexált kódolást használja. E szabványok jellemzője, hogy nem kizárólag időben, hanem a nézőpontok között viszonyítva térben is alkalmazza az MPEG szabványban már bemutatott redundancia keresési eljárásokat.

A 3D kódolás szabványosításának kezdetét 1996-tól, az MPEG-2 MultiView Profile szabvány megjelenésétől számíthatjuk. A H.264/MPEG-4 AVC (eredeti) szabványnak a több nézőpontú megjelenítés támogatása még nem volt része, de kiegészítették 2009-ben a Stereo High Profile-al, ami jelentős előrelépés a szabványban. Egy fontos kiegészítése a szabványnak a Frame Packaging Arrangement SEI, ami kiegészítő információkat rögzít a kódolt jelfolyamról. Ennek a Stereo SEI a mostanában használt része, amely a dekódoláshoz információkat ad az egyes képkockáknak a sztereó-elrendezéséről, hogy például váltott oszlopos, váltott soros, vagy sakktáblaszerűen elrendezve tartalmazza a bal és jobb szem számára az alkotandó képnek az elemeit [22].

Közös jellemzője a mélységalapú formátumoknak, hogy a 2D képadatokat (amelyek a jelenlegi infrastruktúrával is értelmezhetőek) kiegészíti a rögzített jelenet objektum-pontjainak kamerától számított mélységadataival, mely egy csatolt adatfolyamban szinte egy külön képként tárolható. Az MPEG-C Part 3 az egyik meghatározója e megközelítésnek. A megoldás előnye, hogy a mai széles körben elterjedt infrastruktúrával teljes mértékben kompatibilis (a 2D képadat-folyam lehet MPEG-2 valamint H.264/MPEG-4 AVC szabványú is). További előnye, hogy a mélységtérképre támaszkodva számolhatók a több nézőponthoz (ilyen megjelenítéshez) szükséges sztereoszkopikus képadatok is, valamint előny az is, hogy a megjelenítéskor állítható a mélységtartomány kompressziója is.

A H.264/MPEG4 AVC érdekessége, hogy már tartalmaz egy (még) nem szabványosított kiegészítést is. Ez a több nézőpontú sztereoszkopikus megjelenítés kódolására vonatkozik. A több nézőpontú kódolás egyik kézenfekvő megközelítése,

hogy minden nézőpontot külön-külön adatfolyamként tárolunk. Ennek ugyan előnye, hogy alacsony a számításigénye, ugyanakkor nagyon sok benne a redundancia, mind térben, mind időben. Az említett szabvány kiegészítés az előbbiek szerinti redundanciát csökkenti azzal, hogy figyelembe veszi a térbeli és időbeli összefüggéseket is.

[22] az MPEG-4 AFX (Animation Framework eXtension) kiegészítéséről azt emeli ki, hogy ez a pontfelhő, a mélységtérkép és a nézőpontfüggő multitextúrázás alapú megjelenítés integrálására törekszik. Az MPEG-4 AFX DIBR már a rétegzett mélységtérképek használatán alapul. Ez több nézőpontból is rögzített képet tartalmaz, mindegyik képhez saját mélységtérképpel. Ez lehetővé teszi az előbbiek szerinti adatokból a köztes nézőpontoknak megfelelő képek számítását is. Érdeemes megemlíteni még a pontfelhő-alapú megközelítést, amely a kép minden pontjának 3D koordinátáját és felszín-adatát rögzíti. Ennél minden tárolt pont tartalmazza azokat az attribútumokat, amelyek a megjelenítéshez szükségesek, ezáltal ez egyszerűsíti a 3D megjelenítést.

2.4. Képmínőség és mérése

Egy kép minőségét említve számos különböző dolog juthat eszünkbe. Gondolhatunk a tartalmi megközelítésekre, egy fotó esetén például a téma elhelyezésére, vagy akár a helyes expozíciós idő eltalálására – egy alul vagy túlexponált képet illethetünk a rossz minőségű jelzővel. Gondolhatunk ugyanakkor a kép fizikai megjelenésére is, a felbontásra, a kontrasztviszonyokra, a színtelítettségre. A képmínőség fogalom értelmezésénél még nagyobb nehézséget jelenthet, hogy a szemlélők véleménye időben is változhat, vagyis egy kép minőségéről alkotott vélemény függhet a megtekintő előképzettségétől, hangulatától vagy éppen a megtekintés céljától. [3]

2.4.1. Képmínőség értelmezése

A képmínőség tehát egy erősen szubjektív fogalom, függ a kép tartalmától, a keletkezés körülményeitől, a megjelenítés módjától, a megtekintés céljától, és a megtekintő állapotától egyaránt.



2.5. ábra. Tartalmilag, illetve jelalak torzulás miatt rossz minőségűnek minősített képek

Értelmezhetjük a képminőséget esztétikai illetve tartalmi szempontok szerint – a 2.5. ábra bal oldali képét nem neveznénk jó minőségűnek, mivel a téma takarásba került. A képminőség jelentheti ugyanakkor a kép jelalakjában bekövetkezett torzulás mértékét is –míg az említett ábra bal oldali (tartalmilag rossz minőségű) képe jelalak szempontjából nem szenvedett torzulást, látható, hogy a jobb oldali kép képen JPEG kódolás okozta torzulásai már kép esztétikai értékelését is elronthatják.

A digitális képfeldolgozásban legtöbbször a jelalak változásával, illetve az annak torzulása következtében felmerülő esztétikai minősítéssel foglalkozunk, a tartalmi szempontok nem képezik a képminőség mérés tárgyát. Legtöbbször azonban nem elégséges a képminőség nehezen összehasonlítható, szubjektív értékelése – például az egyes veszteséges kódolók kimeneteinek értékelésénél, vagy az adattovábbítás során keletkezett hibák, illetve zajhatások súlyosságának mérésénél objektív megközelítésre van szükség. Vagyis a képminőséget, mint objektív, mérhető jelenséget kell értelmezni, mely csak a kép jelalakjától függ. E megközelítés előnye, hogy megismételhető módon teszi lehetővé összehasonlítását, illetve rangsorolását minőség szempontjából.

2.4.2. Képminőség objektív és szubjektív mérése

A képi minőség objektív mérése a kép jelalakja alapján megfelelő algoritmus segítségével történhet, melynek kimenete egy a kép minőségét jellemző szám. Ilyen

algoritmus, vagy módszer előállításához azonban először is szükség van a képek minőségének egy szubjektív vizsgálatára, mivel a gépi minőség értékelést az emberi minőség értékeléshez kell a lehető legjobban közelíteni.

A képi minőség szubjektív mérésének kérdéseivel az ITU-R BT.500 ajánlás foglalkozik [28]. Az ajánlásban több minőség értékelő, illetve rangsoroló teszt leírása szerepel, elsősorban televízión megjelenített képekre – szerepelnek összehasonlító tesztek, minőségi rangsorolások, illetve szubjektív minőség osztályzások is. Az ajánlásban megfogalmazott minőség mérő eljárások kimenete az egyes képekre egy minőségi pontszám, vagy képek egy rangsorolása. Az ajánlás a vizsgálatok pontos lefolyására, a megjelenítők kritériumaira, a résztvevők számára, a megfigyelés körülményeire és idejére egyaránt részletes leírást tartalmaz.

Egy objektív képminőség mérő jelzőszám kifejlesztése, módosítása általában egy-egy szubjektív kísérletsorozaton alapul, a jelzőszám annál jobban használható, minél jobban közelíti a szubjektív értékelés eredményeit.

2.4.3. Képminőséget jellemző jelzőszámok

A képi minőséget jellemző jelzőszámok általában egy kép torzultságát mérik, illetve próbálják kifejezni, például egy „tökéletes” torzulásmentes (referencia) képhez képest. A jelzőszámokat többféleképpen is csoportokra oszthatjuk. Az alapján, hogy igénylenek-e referencia képet, a következő három nagy csoport különíthető el: teljes-referencia alapú módszerek – a torzulást a referencia képhez képest adják meg; részleges-referencia alapú módszerek – az eredeti jelre való utalásokat kiegészítő információként kapják meg; nem-referencia alapú módszerek – nem rendelkeznek az eredeti jelre vonatkozó információkkal, egy kép minőségét kizárólag jelalakja és jellemzői alapján határozzák meg [23].

Másrészt csoportosíthatjuk a módszereket az alapján is, hogy figyelembe veszik-e a kép-specifikusságot (vagyis hogy éppen képeket és nem például elektromos jeleket hasonlítanak össze). Ezen szempont alapján megkülönböztethetjük a klasszikus jelfeldolgozási módszereket, illetve az emberi látórendszert figyelembe vevő kép-specifikus módszereket. A következőkben ismertetett jelzőszámok eredményeit, illetve használhatóságát a 2.6. ábrán szereplő példán keresztül szeretném

bemutatni. Az „a” jelű képen az eredeti torzítatlannak elfogadott referencia képet láthatjuk, a további képek az alábbi torzításokon estek keresztül:

- a) referencia kép, nincs torzítás
- b) az egész kép eltolása felfele 5 pixellel, a felszabadult rész fehér színnel kitöltve
- c) zaj hozzáadása az ún. HVS (Hue-Saturation-Value) színtérben a képhez
- d) felirat hozzáadása a képhez
- e) a kép torzítása RGB zajjal
- f) a kép elmosása Gauss-szűrő használatával



2.6. ábra. A tihanyi apátság fotója, és torzított változatai

A képek minőségi sorrendjének meghatározása nem egyértelmű, azt szubjektív szempontok befolyásolják. Azonban mindenképpen fontos kérdés, hogy az objektív módszerek mennyire közelítenek a szubjektív rangsorhoz, vagyis az emberek

véleményéhez. Az általam feljebb megadott képsorrend, „a”-„f” saját szubjektív minőség rangsoromnak is tekinthető, a továbbiakban ezt fogom referenciának használni.

2.4.3.1. Klasszikus jelfeldolgozási módszerek

A képek, mint digitális jelek minőségét lehet a jelfeldolgozásban elterjedt módszerek segítségével összehasonlítani, és ezek segítségével jelzőszámokat alkotni. Ezen összehasonlító módszerek általában két képet (jelet) dolgoznak fel, és a köztük lévő különbség mértékét jellemzik, vagyis teljes-referencia módszereknek tekinthetők.



2.7. ábra. Az MSE eredményei a tesztképekre

MSE A *Mean Square Error* vagyis Átlagos négyzetes hiba módszer az egyik legelterjedtebb módszer két jel összehasonlítására, használata és számítása a statisztikai szórásnégyzethez hasonlatos. Az MSE számításának módját a 2.1. képlet

adja meg, ahol f és g a két összehasonlítandó kép két dimenziós mátrixként reprezentálva, N és M a kép magasságának, illetve szélességének jelölésére szolgál.

$$\text{MSE}(f, g) = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M (f_{i,j} - g_{i,j})^2 \quad (2.1)$$

Az MSE az egyik legszélesebb körben használatos eszköz a jelfeldolgozásban, mivel egyszerű, könnyen számítható. Képek minőségének meghatározására azonban csak korlátozottan használható.

A 2.7. ábrán láthatjuk az MSE algoritmus „rangsorát” a 2.6. ábrán bemutatott tesztkép sorozatra (a betűjelek az előző pontban bemutatottaknak megfelelően). Láthatjuk, hogy mivel az elmosás a pixelek értékét csak kicsit befolyásolja, így a négyzetes hiba kicsi lesz, vagyis az MSE jónak értékeli azt a képet, amit szubjektíven a legrosszabbak közé soroltam, míg a pár pixeles eltolást (melyet alig észlelünk), a legnagyobb eltérésűnek értékeli.

PSNR A *Peak signal-to-noise ratio* vagyis a csúcs jel zaj viszony az MSE egyfajta továbbfejlesztése, mely a jel lehetséges maximális erősségének, és az azt terhelő zaj intenzitásának összevetéséből számítható ki a 2.2 képlettel – ahol MAX jelöli az adott kép pixeleinek maximális intenzitását. A PSNR szintén a teljes-referencia (*full reference*) módszerek közé tartozik, vagyis egy képet a referencia képpel hasonlít össze.

$$\text{PSNR}(f, g) = 10 \cdot \log_{10} \left(\frac{\text{MAX}_f^2}{\text{MSE}(f, g)} \right) \quad (2.2)$$

A PSNR előnye talán az MSE-vel szemben a nemlineáris jelleg, ami valamennyire megfelel a látásmechanizmusnak, de megemlítenéd, hogy a fenti példán mutatott jelenség alapvetően itt is fennáll, tehát a szubjektív rangsor nagyban különbözhet a PSNR által adott eredménytől.

2.4.3.2. HVS-en alapuló jelzőszámok

Látható, hogy a pusztán jelalakokon és különbségükön alapuló jelzőszámok értékelése távol áll az emberek szubjektív véleményétől, ez meglehetősen korlátozza ezen jelzőszámok alkalmazásának lehetőségeit. A HVS (*Human visual system*), azaz az

emberi látórendszer viselkedésének értelmezése tehát a feladat. Olyan jelzőszámokat kell kidolgozni, melyek a lehető legjobban közelítik a HVS tulajdonságait, azaz az emberi látórendszer sajátosságait (2.1.2. pont).

SSIM A *Structural SIMilarity index* [42] vagyis strukturális hasonlósági szám a fénysűrűség és a kép struktúrájának figyelembe vételével értékeli a kép minőségét. A módszert a 2.3. képlet adja meg, ahol μ az adott kép átlagát, σ a kép varianciáját, σ_{xy} a két kép kovarianciáját jelöli, c_1 és c_2 a kép fizikai reprezentációjából adódó konstansok az osztás stabilizálására. ($c_1 = (k_1 L)$, $c_2 = (k_2 L)$, ahol $k_1 = 0.01$, $k_2 = 0.03$ alapértelmezett esetben, L pedig a képdinamika tartománya.)

$$\text{SSIM}(f, g) = \frac{(2\mu_f\mu_g + c_1)(2\sigma_{fg} + c_2)}{(\mu_f^2 + \mu_g^2 + c_1)(\sigma_f^2 + \sigma_g^2 + c_2)} \quad (2.3)$$



2.8. ábra. Az SSIM eredményei a tesztképekre

Az SSIM módszerrel a tesztképen kapott eredményeket a 2.8. ábra szemlélteti. Látható, hogy az előző módszerekhez képest jobban közelíti a szubjektív rangsort. Többféle továbbfejlesztése létezik, ilyen például a MS-SSIM (*Multi-Scale Structural SIMilarity*), mely az SSIM algoritmusát többször, különböző felbontású képváltozatokon futtatja le [45]. A következőekben bemutatott képsúlyozási eljárásokkal együtt is sokszor alkalmazzák [43].

2.4.3.3. Pooling technikák

Az eddig bemutatott módszerek a teljes képet azonos súllyal veszik figyelembe, függetlenül attól, hogy egy adott torzulás a kép mely részén következett be. (Az SSIM már némi képjellemző alapú megkülönböztetést tartalmaz.) Az ember azonban nem egyenlő figyelmet fordít a kép teljes területére, egyes területek az ember számára fontosabbak, míg más területek elhanyagolhatóak - valamint a látás sajátosságainak köszönhetően különböző területeken a hibaérzékenységi küszöb is más lehet. A pooling technikák egy-egy stratégián alapulva megpróbálják kiválasztani és súlyozni a kép egyes területeit úgy, hogy az ember számára fontos, vagy hibaérzékenyebb területek magasabb súllyal szerepeljenek.

Minkowski pooling

$$Q = \frac{1}{N} \sum_i^N q_i^p \quad (2.4)$$

A Minkowski pooling a 2.4. képlet alapján számítható ki, ahol q_i az i .pozíción számított lokális torzulás értéke, p pedig a Minkowski kitevő. A kitevő értékének növelésével az erősen torzult helyek nagyobb súllyal szerepelnek a kép értékelésében, mivel kimutatható, hogy az erősen torzult területek jobban befolyásolják a hiba észlelését, így kis területen erős torzulás esetén a képet rosszabb minőségűként észleli [44].

Saliency alapú pooling Saliency kifejezés alatt a képek olyan általában alacsony szintű, kiugró tulajdonságait értjük, melyek megragadják az ember figyelmét, ezáltal a területek fontossá válnak az ember számára. A saliency alapú súlyozás ezen sajátosságos pontokon, területeken alapul, melyek egy súlytérképen nagyobb súlyt kapnak – az ilyen területen bekövetkezett torzulás tehát nagyobb

súllyal szerepel az összértékelésben.

$$Q = \frac{\sum_i^N w_i q_i}{\sum_i^N w_i} \quad (2.5)$$

Számítása a 2.5. képletben alapul, ahol q_i az i . pozíción számított lokális torzulás értéke, w_i pedig az adott pozíció, képterület súlya. A súlytérkép előállításához rengeteg képtulajdonság áll rendelkezésre, a lokális változástól kezdve a kontraszton át az éltávolságig [43].

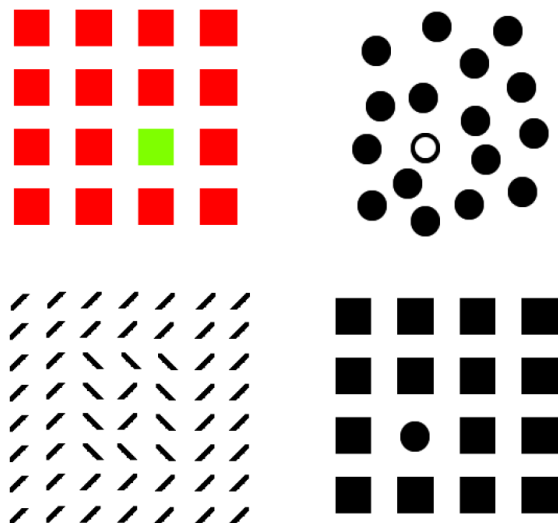
Objektum alapú pooling Az objektum alapú pooling számítási módja megegyezik az előző pontban bemutatott saliency alapú technikával, egyetlen különbség, hogy jelen esetben a súlytérkép előállításához nem a kép alacsony szintű tulajdonságait, hanem a magas szintű feldolgozás eredményeit vesszük figyelembe. A súlytérképet ilyen esetben általában egy képszegmentáló, objektumfelismerő eljárás adja. A képen található bizonyos objektumok, pl. arcok, szövegek ugyanis megragadják az ember tekintetét, torzulásuk esetén a kép hamar élvezhetetlenné, rossz minőség érzetűvé válik, ezáltal szükséges e területek fokozott figyelembe vétele.

Új módszerek Az előzőekben bemutatott képminőség mérési, súlyozási módszerek a képminőség észlelés egy-egy sarkalatos szempontjára koncentrálnak. Az eljárások különböző kombinációival a szubjektív értékeléshez még közelebbi eredmény érhető el, illetve további pooling módszerek találhatók, ilyen például egy a referencia, és torzult kép közötti kölcsönös információn alapuló pooling technika és az MS-SSIM kombinációjaként kialakított IW-SSIM algoritmus [43].

2.5. Kiugró pontok („saliency”) és szerepük a képfeldolgozásban

A képi információ feldolgozása két némileg ellentétes folyamat eredménye. Az egyik folyamat kifejezetten a képi információn alapul, a szem által érzékelt elektromágneses sugárzásból indul ki – ezt *bottom-up* feldolgozásnak nevezzük [24]. A *top-down* folyamat ellentétben általában szemantikai szempontokon alapul, a kép

tartalmi feldolgozásától függ. A folyamatokat alacsony, illetve magas szintű feldolgozásnak is nevezzük. Magas szintű vagy top-down feldolgozás dominál például keresési feladatok esetén, mely esetben elsősorban az objektumok, illetve a kép szemantikai tartalma kap hangsúlyt az észlelés során.



2.9. ábra. Példák saliency régiókra. [17]

Látható, hogy a képminőség értékelésénél fontos tényező lehet az adott képen a fontos képrészletek helyzete. Az alacsony szintű folyamatok kiindulása a képi információ, ahol kiugró pontok ragadhatják meg a tekintetet. Ezeket képrészleteket nevezzük *saliency* pontoknak. Kiugró pontok lehetnek eltérő színű, alakú, irányítású, mozgású objektumok, objektum-részletek, élek, sarkok – klasszikus irodalmi példákat láthatunk a saliency jelenségére a 2.9. ábrán. Jellemző pontok detektálása ma is kutatott terület (akár a képminőség mérési kutatásokban is), alacsony szinten kereshetők él-, illetve sarokdetektorokkal. A kiugró pontok különböző skálákon is megjelenhetnek – különböző nagyságú képrészletek, akár teljes objektumok is lehetnek, ilyen tulajdonságú pontok akár SIFT illetve SURF leírókkal is detektálhatók [20, 15].

Az objektív képminőség értékelő jelzőszámok kialakításánál a cél olyan saliency detektáló algoritmus kifejlesztése, mely által kijelölt fontos képterületek egybeesnek a (kísérletileg meghatározott) valós fixációs pontok helyzetével. Következő-

ekben bemutatott kutatásunkban saliency detektáló algoritmusokat nem használtunk, kizárólag a fixációs pontok helyzetével foglalkoztunk.

3. Sztereó képekkel végzett vizsgálatok

3.1. Célkitűzés

Napjainkban nagy sebességgel nő az értékesített 3D megjelenítők aránya, s ezen folyamat erősödését jósolják [2], így előtérbe kerülhetnek a sztereó látás során tapasztalt képminőség észlelés kérdései. Korábbi megfigyeléseink során a sztereó képeknél a hagyományos változatoktól eltérő, szokatlan észlelési aspektusokat tapasztaltunk, így érdemes vizsgálatokat folytatni a sztereó képminőség észlelés sajátosságainak feltárására.

A képek kódolása esetén általában korlátozott sávszélesség áll rendelkezésre. A jobb minőség több adatot, a kevesebb adat jelentősebb képi torzulásokat jelent, így meg kell találni a megfelelő egyensúlyt a képminőség és a bitráta között. Célunk nem a jel lehető legkisebb változásának, hanem az ember által legkevésbé észlelhető torzulásának előállítás. A 3.1. ábrán látható képeket ugyanolyan mértékben, különböző helyeken torzítottam. Látható, hogy ugyanakkora torzulás a kép más-más részein különböző minőségi változást okoz. Konceptiónk szerint a képminőség érzet javításához meg kell ismernünk az emberi látórendszer mechanizmusait, sztereó megjelenítés esetén mutatott sajátosságait, ez alapján következtethetünk az ember számára fontos képjellemzőkre.

A képi minőség javítására két mód áll rendelkezésünkre. Megismerjük az ember számára fontos képterületeket és ezeket próbáljuk jó minőségben kódolni, vagy azonosítjuk azon területeket melyeken az emberi látórendszer ingerküszöbe magas, vagyis kevésbé érzékeny a torzulásokra, ami által sávszélességet spórolhatunk a kép többi területének. Mindkét módszer a sávszélesség egyenlőtlen elosztásán alapul



3.1. ábra. A két kép ugyanolyan mértékű torzulást tartalmaz, mégsem így érezzük

és kulcsa a képjellemzők alapján olyan súlyozás megállapítása, mely leginkább alkalmazkodik az emberi látórendszer sajátosságaihoz.

Az ember képminőség érzetet több tényező befolyásolja, ezek közül a legfontosabbnak tekinthetők:

- hova nézünk a képen – melyek azok a képterületek, ahol a látás fixációi, rövid megállásai, azaz a tényleges információ felvétel fókuszai kerülnek
- mennyire vagyunk érzékenyek – az emberi látórendszer ezen a képterületen milyen érzékenységgel rendelkezik (pl. Weber törvény)
- mekkora a torzulás – az adott képterületen mekkora a torzulás mértéke

Fontos kérdés tehát, hogy a két dimeziós képekhez kidolgozott kódolási, illetve objektív képminőség mérési eljárások lényeges változtatás nélkül alkalmazhatóak-e sztereó képekre. Az alkalmazhatóság egyik feltételének tekinthető, hogy az ember számára fókuszba kerülő területeken, azaz a fixációs pontok elhelyezkedésén a képmegjelenítés változása ne változtasson. Kutatásunk során célunk a sztereó megjelenítés sajátosságainak vizsgálata olyan súlyozási módszer kidolgozása szempontjából, mely a jelenleg használtaknál (melyek a sztereó tulajdonságokat és disz-

paritás információkat nem veszik figyelembe) jobban közelítik az emberi minőség észlelés sajátosságait.

Kutatásunk jelen első fázisában célunk két megjelenítési mód (hagyományos két dimenziós – monó, és binokuláris távolság észlelés alapú – sztereó) szemlélése közötti különbség keresése, és bizonyítása az ember számára fontostként azonosított képterületek és képjellemzők szempontjából, a fixációs pontok eloszlásának vizsgálatán keresztül. A feladat megvalósítása vizuális teszt-sorozat segítségével történik, mely során szemkövetővel rögzítjük a résztvevők fixációs pontjait, így képet kaphatunk az ember számára fontosnak ítélt képrészletekről. A teszt során rögzített monó illetve sztereó mintákban statisztikai elemzéssel keresünk különbségeket a fixációs pontok eloszlásában.

Kutatásunk folytatásaként a kimutatott különbség okainak feltárására, illetve az eredmények alkalmazására a sztereó képminőséget jellemző módszerek területén további tesztek megvalósítását tervezzük.

3.2. Irodalmi áttekintés

A képminőségi megállapítások és vizsgálatok tekintetében a pszicho-vizuális tesztek alkalmazása az irodalomban széles körben elfogadottnak, és bizonyító erejűnek tekinthetőek. A tesztek általában a terület egy-egy kis szegmensének vizsgálatára végzik el. Két alapvető tesztformát különböztethetünk meg: az egyik esetben egy feladat végrehajtását, másik esetben a kép szabad bejárását vizsgálják (feladat nélkül). Feladat lehet egy kísérletnél például egy 3D-s térben található pont megérintése (sztereó képpár információi alapján) [47], vagy képek minőségének osztályozása, illetve összehasonlítása [31]. Ezeknél a teszteknel a feladat végrehajtásának sikerességét, módját, sebességét vizsgálják, nincs szükség a szem követésére. A szabad bejárásos, feladat nélküli kísérletek esetében azonban szükséges a szem mozgásának követése, a fontos területek megállapításához. Számos kísérletet végeztek ilyen módon is, például generált 3D objektumokon fixációs pontok jellemző helyének megállapítására [18], vagy abnormális színi hatások vizsgálatára [9]. E cikk megállapításai közé tartozik, hogy a színek nem befolyásolják a fixációs pontok helyét, azok abnormális színek, illetve a színek elvétele esetén is közel azonos helyen azonosíthatóak, azonban az objektumok felismerését a színek hiá-

nya lassíthatja. Ez alapján elmondhatjuk, hogy a vizuális érzékelést, a közvetlen szemmozgást nem befolyásolják ezek a momentumok, azonban az agyban kialakult hatásokat, azaz az észleletet lassíthatják vagy befolyásolhatják is, az érzet és az észlelés különválasztására néhány esetben én is támaszkodtam kutatásom során.

A sztereó képtömörítés módszereit vizsgálja szubjektív értékelésen alapuló vizuális teszt segítségével [30]. A kísérletben mindkét szemnek szánt képet JPEG kódolással, különböző mértékben tömörítik szimmetrikusan, illetve asszimmetrikusan is (az egyik kép más minőségű) – a kísérlet a monó és sztereó eredmények között lényeges különbséget nem állapít meg, de megállapítja, hogy az eredmények meglehetősen képtartalom függőnek tekinthetők.

Néhány esetben találhatunk ellentmondást is az irodalmi eredmények között. Az arc felismerésének összehasonlításának vizsgálata során a monó illetve a sztereó észlelések között nem állapított meg különbséget [11]. Egy későbbi vizsgálat alapján azonban az arcfelismerés sebességének és pontosságának növekedését állapították meg sztereó esetben [4], vagyis az irodalmi eredmények, azaz a vizuális tesztek eredményei nem mindig egyértelműek, azok a feladat megfogalmazásától illetve a felhasznált mintáktól (képek) egyaránt szorosan függnak. A téma irodalmának feldolgozása során, a mi célkitűzésünkhöz hasonló jellegű összehasonlító kísérletet nem találtam, azaz tudomásom szerint a fixációs pontok eloszlásának különbségét mások nem vizsgálták.

3.3. Előzmények és korábbi tapasztalatok

3.3.1. Minőségromlás hatása a térérzetre

Pszichológiai kutatások bizonyítják, hogy az agy rendkívüli módon tud alkalmazkodni a vizuális ingerek változásaihoz. Egyes kísérletekben függőlegesen tükröző lencsét adtak a kísérlet résztvevőire (tehát mindent fejjel lefelé láttak), s volt, aki 7. napon már közlekedett, 14. napon síelt [21]. Az agy tehát képes megszokni, illetve korrigálni a szemből felé érkező fals információkat.

Informatikai megközelítés szerint a képek és videók átvitelénél, megjelenítésénél a csatorna kapacitásra is optimalizálni kell, ezáltal adott sávszélességen a

lehető legkevesebb adatot próbáljuk átküldeni, ez vezetett a képek tömörítésének kérdéséhez. A képek veszteséges tömörítését azonban úgy kell elvégezni, hogy az eredmény lehetőleg jóval kisebb adathalmazt eredményezzen (ezáltal erőforrásokat tudunk spórolni), azonban a szemlélőt ez a lehető legkisebb mértékben zavarja. Érdekes vizsgálnunk tehát, hogy az emberi látás mennyire érzékeny a különböző szándékos, tömörítési célú minőségrontásra, illetve az esetleges átvitel közben keletkezett torzulásokra. Jogosan merülhet hát fel a kérdés, hogy 3D-s képek esetén hogyan reagál az agy az egyik, vagy mindkét kép torzulásaira.

3.3.1.1. Elmosás hatása a térérzetre

Saját megfigyeléseim [12] és irodalmi példák [47, 46] is alátámasztják, hogy az emberi látórendszer meglehetősen érzéketlen az egyik kép elmosódására, a hiányzó magas-frekvenciás információkat a másik szem képéből kipótolja. Az egyik szem képének erős elmosódása esetén is tehát a térélmény, illetve a kép minőségének értékelése is egyaránt jó lehet – sőt a 3D-s céltárgy lokalizálási pontossága esetében megfigyelhető egy érdekes jelenség: a két kép egyikének elmosásával javul a vizuális teljesítmény [47]. Ennek egyik lehetséges magyarázata, hogy a sztereopszis esetében az egyik szem „vezető” szerepet tölt be, és éles képeket küld az agynak, míg a másik szem csak a parallaxist (elérést) méri a térhatás-észleléshez (3D látáshoz). A két kép egyikének mérsékelt homályosodása megkönnyítheti az agynak a „vezető” szem kiválasztását.

3.3.1.2. Hibaérzékenység a magas frekvenciás torzulásokra

Mivel a szem a magas frekvenciás jelek hiányát a másik szem képéből pótolja, felmerülhet a kérdés, mi történik, hogyha tömörítés hatásainak következtében (blokkosodás – pl. JPEG tömörítés), vagy átviteli hiba, termikus zaj következtében („salt and pepper” zajhatás) magas-frekvenciás jelek adódnak a sztereó képpár egyik képéhez. Megfigyeléseink alapján [12] a magas-frekvenciás hatások nagyobb mértékben zavarják a szemlélőt, mint az alacsony frekvenciás torzulások, de a térélményt ezek sem nagy mértékben befolyásolják.

3.3.2. Következtetések

Korábban megfigyeltük tehát, hogy a sztereó módon megtekintett képek esetében, a normál képekre érvényes, és bevált megfigyelések nem feltétlenül használhatóak maradéktalanul. Egy speciális, technikai jellegű, de szemléletes példa lehet, hogy anaglif képek kódolására a JPEG tömörítés meglehetősen alkalmatlan. A JPEG által alkalmazott színtér-transzformáció miatt ugyanis az RGB színcsatornák két külön képre leválogatott információit „összekeverik”, ezáltal az elkészült sztereó képen erős szellemképek jelennek meg, vagyis a két kép szétválasztása nem biztosított.

A sztereó képek hasonló sajátosságai miatt a monó képek esetében használatos képminőséget jellemző HVS alapú jelzőszámok sem biztos, hogy módosítás nélkül átvihetők sztereó esetre.

4. Vizuális tesztek

A sztereó képek észlelési sajátosságainak vizsgálatához első lépésként a következő kérdés megválaszolására van szükség: találhatunk-e különbséget a monó, illetve sztereó módon megtekintett képek „nézési módja”, észlelési folyamata között, azaz mást, vagy másképp nézünk-e egy három-dimenziós, sztereó képen, annak monó változatához képest. A kérdés megválaszolása vizuális tesztek segítségével lehetséges, melyek során a résztvevők szemmozgását nyomon követve jellemezhetjük a képek megtekintésének folyamatát, ezzel objektív képet kapva a fókuszba kerülő képterületekről, azaz a fixációs pontokról.

Munkánk során két tesztet állítottunk össze és végeztünk el, amelyekben a résztvevők természetes, illetve mesterségesen generált képeket tekintettek meg feladat nélkül, miközben szemmozgásukat követve a képek bejárását (fixációs pontok helyét) rögzítettük.

4.1. Vizuális tesztek előkészítése

A teszt lebonyolításához egy Leonar3Do sztereó megjelenítő eszköz, illetve egy pupilla felismerést és követést támogató szemüveg és szoftver állt rendelkezésre, melyeket a későbbiekben részletesen bemutatok (4.2. pont). A teszt összeállítása során az ITU vizuális tesztekre vonatkozó ajánlásából [28], illetve irodalmi példák-ból indultunk ki [30, 11, 4]. A teszt összeállítása során meg kellett határozni a képi adatbázis összetételét, a megjelenítés módját és idejét, a monó sztereó váltások gyakoriságát, az optimális megtekintési távolságot. A vizuális teszt összeállításához több kísérletet és próba mérést végeztünk [13]. A rendelkezésre álló pupilla követő alkalmazásához emellett gondoskodni kellett a fej megfelelő „rögzítéséről”, megtámasztásáról is.

Az első kísérleteket áll-támasszal végeztük, 110 db természetes képpel, a teszt teljes ideje (kb. 30 perc) így az ITU ajánlásának felső határára esett. Megállapítható volt, hogy ilyen hosszúságú teszt során nem megfelelően biztosítható a fej mozdulatlansága, így a képsorozatot rövidebb részekre osztottuk, valamint a képi adatbázis méretét is jelentősen csökkentettük (40 db kép) – a második tesztben még kevesebb képet (22 db) alkalmaztunk. A fej megtámasztásra is új, a fejt hátulról támasztó eszközt készítettünk, mellyel az esetleges elmozdulás mértéke csökkenthető volt, illetve biztosítható maradt az operátornak szükség esetén adható reakció lehetősége is.

Az optimális megjelenítő távolság (megtekintési távolság) kiválasztása is nehéz feladatnak bizonyult. A megtekintési távolság növelésével ugyanis a mélységérzet is változik, ugyanakkor a szemmozgás is jelentősen csökken, ami a pupilla követés pontosságának csökkenésével járhat (ugyan akkora képterületet a szem jóval kisebb ugrásokkal jár be). A megtekintési távolság csökkentése esetén azonban megnő a bemozdulás, a fej elmozdításának valószínűsége, mivel a perifériás területek bejárásához a túl nagy szemmozgás helyett az ember ösztönösen a fej elfordításával reagál – a fej elmozdulása esetén (kalibrált helyzetből való kimozdulás) a szemüveg helyzete megváltozik, így a pupilla helyzetét nem lehet pontosan megfeleltetni a képernyő egy területének. Az optimális megtekintési távolságot végül az irodalmi példáknál némileg nagyobbban 1.5 m-nek állapítottuk meg, az egyes képek vetítési ideje 15 sec lett.

A második teszt előkészítése, képanyagának kiválasztása az első teszt eredményei alapján történt – a kísérlet technikai lemenetén, beállításain lényegesen nem változtattunk. Mivel az első teszt során alkalmazott természetes képeknél a mélységtérképek megfelelő minőségű előállítására nem volt lehetséges, így a második kísérlethez ismert mélységtérképű laboratóriumi természetes képeket, illetve mesterségesen generált grafikákat használtunk fel.

4.2. Felhasznált eszközök

A vizuális tesztek előkészítéséhez és végrehajtásához az alábbi eszközök álltak rendelkezésre:

- Leonar3Do [19] hardver, szoftvercsomag: Aktív folyadékkristályos, sztereó megjelenítést segítő eszköz, mely biztosítja az LCD szűrős szemüveg folyamatos szinkronizációját a monitorral, ezáltal sztereó módban a jobb és bal szem különböző képeket láthat.
- Samsung SyncMasters 2233 monitor: 120 Hz képfrissítési frekvenciájú monitor sztereó alkalmazásokhoz. A magas képfrissítési frekvencia biztosítja, hogy külön-külön a jobb és bal szem is villogásmentesnek érzékelt képet láthasson.
- ArringtonResearch ViewPoint GigE-60 szemkövető: hardver és szoftver, a szem mozgását kamerával figyeli, kalibráció után a megtekintett monitor pozícióról szolgáltat koordináta információkat. A szoftverrel az adatokon előfeldolgozás (pl. az adatok elő-szűrése, közeli pontok összevonása) végezhető.

4.2.1. Sztereó megjelenítés

A kísérletekhez az aktív folyadékkristályos szemüveges sztereó megjelenítési technikát választottuk, mely szín-torzítás nélkül képes sztereó megjelenítésre (eltekintve a kicsit sötétebb képtől). A kísérletekhez egy sztereó megjelenítő alkalmazást készítettem, mely alkalmas egy előre meghatározott képszekvencia diavetítés-ként való lejátszására, akár monó akár sztereó módban.

4.2.2. Pupilla-követés



4.1. ábra. A teszteknel használt összeerősített szemüvegek

A szemkövetést (pupilla-követést) a ViewPoint GigE-60 szemkövető eszköz segítségével végezzük. Az eszköz egy szemüveg, mely egy kis kamerával a szem képét figyeli (a szemet nem látható infra tartományban megvilágítja a jobb felvétel érdekében), és a képet rögzíti. Kísérleteink során a résztvevők a sztereó megjelenítéshez már viseltek egy szemüveget, így a szemkövető szemüvegét arra erősítve használtuk. Az összeillesztett szemüvegeket a 4.1.ábra szemlélteti. A szemkövető szoftvere a szemüveg kamerája által rögzített kép (a képen pupilla és mozgásának detekciója), és előzetes kalibráció alapján meg tudja állapítani, hogy viselője hova nézett az adott pillanatban. A hardver-szoftver csomag kimenete a szemmozgásra jellemző fixációk és szakkádok sorozata egy fájlban rögzítve.

4.3. A teszt lefolyása

A teszt során a résztvevőknek a megjelenítőtől kb. 1,5 m távolságra fejüket egy speciálisan erre a célra készített fejtámaszon megtámasztva kellett ülni. A résztvevők sztereó látásának ellenőrzése (randompont sztereogramok [16]), valamint a pupilla követő szemüveg kalibrációja után (felvillanó pontokra történő fókuszálási feladat) egy előzetesen meghatározott képszekvencia képeit kellett megtekinteni. Erre azért volt szükség, hogy azokat a személyeket kiszűrjük a kísérletekből, akiknek az átlagostól jelentősen eltérő („gyenge”) a sztereó látása. Első lépésként a feladat, hogy a képek között a képernyő közepén elhelyezkedő kb $1\text{cm} \times 1\text{cm}$ -es szürke négyzetre kellett tekinteni, melyet kalibrációs ellenőrző mintaként használtunk.

Egy résztvevőnek az első tesztben 4, a második tesztben 2 képszekvenciát vetítettünk egymás után. Minden képszekvencia vetítését külön kalibrációs folyamat előzi meg. A képszekvenciák megjelenítési típus szerint homogének, azaz vagy tisztán monó, vagy tisztán sztereó módon megjelenített képeket tartalmaznak, képszekvencián belül monó-sztereó váltás nincs, mivel a sztereó képek megtekintéséhez megfelelő adaptáció szükséges. A teszt során operátor működött közre, aki az eredményeket valós időben nyomon követte, ezáltal súlyos bemozdulás, vagy más hiba esetén a tesztet félbeszakíthatta, a képsorozatot újraindíthatta.

A teszt résztvevői minden képet pontosan egyszer tekintettek meg, vagy monó vagy sztereó módon, azaz ugyanazt a képet mindkét megjelenítési módon egy

részrtvevő sem látta, de minden résztvevő látott mind monó mind sztereó módon megjelenített képszekvenciát.

4.4. Rögzített adatok

A tesztek során a 4.2. pontban bemutatott szemkövető eszköz szoftverével rögzítettük a résztvevők szemmozgását (fixációs pontok, szakkádok) és a képernyőképet 4 képkocka/másodperc mintavételezési sűrűséssel. A szemkövető rendszer által küldött nyers adatokat az eszköz saját szoftverével előfeldolgoztuk, mely során a következő adathalmaz állt elő: a képszekvenciák megtekintését fixációk és szakkádok váltott sorozataként jellemezzük. Minden egyes rekordhoz rögzítjük annak kezdő időpontját, időbeli hosszát és térbeli koordinátáit (szakkádoknál kezdő és végpontra is), melyeket a rendszer 0 és 1 közötti értékként normál két tizedes jegy pontossággal. Meghatározható tehát, hogy mérésünk elvi (kizárólag a kerekítésből adódó) felbontása a 4.1. képlet segítségével adható meg, ahol W a képernyő szélességét, N az ábrázolható számok mennyiségét jelöli.

$$\text{Elvi felbontás} = \frac{W}{N} = \frac{1680}{100} = 16.8\text{pixel} \quad (4.1)$$

Mivel a mérések során a résztvevők bemozdulásából is adódnak hibák (melyeket az adatfeldolgozás során korrigáltunk), így a korrigált adatok felbontása kb. 30-50 pixel közöttire becsülhető, mely értéket az ellenőrzéseink is alátámasztanak.

A rögzített nyers adatok feldolgozása után a képi információ és a rögzített adatok összevetésével egy olyan adatbázis nyerhető, mely minden egyes résztvevő minden egyes fixációs pontjáról tartalmazza annak időbeli és térbeli paramétereit, valamint az adott kép (amelyhez a fixációs pont tartozik) adott környezetének képi tulajdonságait (pl. színtelítettség, lokális kontraszt, éltávolság, diszparitás - második teszt esetén). A feldolgozó szoftverben a rekordok további képi jellemző paraméterekkel szükség esetén bővíthetők.

A két teszt során összesen 47279 rekordot, ezen belül 23627 fixációs pontot rögzítettünk.

4.5. Első vizuális teszt

4.5.1. Képi adatbázis

A tesztek elvégzéséhez változatos témájú természetes sztereó képpárokat gyűjtöttünk internetes forrásokból [33, 34, 35, 36, 37]. A természetes fotók és a változatos témák célja a statisztikai mintákban a tartalomfüggés hatásának csökkentése. A képek felbontása nagyobb volt a képmegjelenítő felbontásánál, így a képeket a képmegjelenítő által megjeleníthető maximális felbontásúra (1680×1050) átméreteztük. A képarányokat megtartottuk, az arány eltérésekből adódó üres felületeket fekete színűnek választottuk, így a 4.2.1. pontban bemutatott programmal a képek teljes képernyőn jelennek meg. A kísérletek során felhasznált összes kép eredetileg sztereó képpár, illetve annak jobb szem számára készült fele – monó megjelenítés használata esetén a jobb szem számára szánt képet jelenítjük meg.

A képeket tartalom alapján alapvetően két csoportba sorolhatjuk. Az egyik csoport egy előtértémát – jellemzően egy alakkal – illetve egy viszonylag semleges monoton háttérrel tartalmaz. A második csoportba olyan képek, kép-párok tartoznak, melyek elő és háttérrel egyaránt mozgalmassá, vagy a képen lineáris távolság hatás figyelhető meg, azaz a kép nem különíthető egyértelműen elő és háttérre, hanem az objektumok nézőtől való távolodása folytonos (4.2. ábra).



4.2. ábra. Példák az első teszt képeiből

A teszt során 40 képet használtunk fel, melyekből nyolc képszekvencia épült fel. A képszekvenciák közül egy résztvevő előre meghatározott négyet látott, melyeket egyenként kalibrációs folyamat előzött meg (egy képszekvencia 10 képet tartalmaz). A képeket tartalom szerint csoportosítottuk – ezáltal az első két kép-

csoport olyan képekből épül fel, melyek közel homogén háttér előtt egy objektumot tartalmaznak, míg a második két szekvencia képeinek háttére is mozgalmasabb, több objektumot tartalmaznak. A kísérlet résztvevőit két csoportba osztottuk („A” és „B”) aszerint, hogy mely képsorozatokot tekintették meg monó illetve melyeket sztereó megjelenítési móddal. Egy résztvevő tehát mind a 40 képet végignézte úgy, hogy egy képet csak egyszer látott vagy monó vagy sztereó módon megjelenítve. A két résztvevő-csoport adatait összevetve tehát minden képről rendelkezünk monó, illetve sztereó adatokkal egyaránt.

4.5.2. Résztvevők

A tesztet kilenc 20 és 35 év közötti egészséges, vagy korrigált látású résztvevővel végeztük el, sztereó látásukat a teszt kezdő fázisában megvizsgáltuk és megfelelőnek találtuk. A kilenc résztvevő közül 5 az „A” jelű képszekvenciákat, 4 a „B” jelű képszekvenciákat láthatta, a résztvevők között 3 nő volt.

4.5.3. Problémák, javítási lehetőségek

Az első teszt során a teszt teljes hossza, a négy képszekvenciával és a hozzá tartozó kalibrációs műveletekkel együtt továbbra is 20-25 perc körül mozgott, mely meglehetősen megerőltető volt a résztvevők számára. A további teszteknel a képek számának csökkentését határoztuk el. Az adatok feldolgozása során kiderült, hogy bár a természetes képeket kellemes sztereó megjelenítővel nézni, azok nem okoznak a szemnek nagy fáradtságot, azonban a nagy felbontású természetes fotókhoz az adatelemzéshez szükséges mélységtérképek, diszparitás térképek előállítása meglehetősen nehézkes - pontos diszparitás térképet a képekhez nem tudtunk előállítani. A további teszteknel így ismert mélységtérképű képek használata szükséges. Ez volt a legfőbb oka, hogy egy második tesztsorozatot állítottunk össze, ahol ismertek a diszparitás térképek.

4.6. Második vizuális teszt

4.6.1. Képi adatbázis

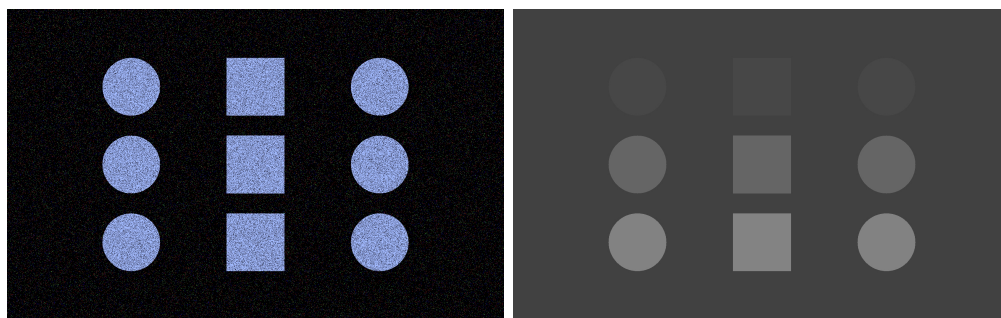
A második tesztnél használt képi adatbázis képei két nagy csoportra oszthatóak. Az első csoportba természetes laboratóriumi fényképek (6 db) sorolhatóak, melyek mélységtérképe ismert (a mélységtérkép nem a két képből utólagosan lett számítva, hanem a felvételkor speciális eljárással határozták meg – ezeket a képeket a mélységtérkép számító algoritmusok tesztelésére használják). A képekhez a ground-truth mélységtérképeket használjuk fel. A második csoportba mesterségesen generált képek kerültek (20 db), melyeken geometriai objektumok, betűk és számok jelennek meg. A mesterséges képek oly módon lettek megszerkesztve, hogy azokon monó megjelenítés esetén semmilyen mélység információ nem jelenik meg (nincs takarás, perspektíva vagy egyéb hatás) – vagyis a mélység érzet sztereó megjelenítés esetén többlet információként jelenik meg.

A második tesztnél felhasznált képek csoportosítása:

- mesterséges képek
 - betűket tartalmazó képek (4 db)
 - számokat tartalmazó képek (4 db)
 - geometriai alakzatokat tartalmazó képek (8 db)
 - randompont sztereogrammok (4 db)
- természetes képek (6 db)

A mesterségesen generált képek nem tartalmaznak teljesen homogén színű felületet, a két kép agyban történő összeecsúsztatásának (regisztrációjának) megkönnyítésére, illetve a szellemkép hatások csökkentésére. A geometriai objektumokat tartalmazó képek, illetve a számokat és betűket megjelenítő képek felének esetében a képenként megjelenített 9 db objektumot rendezett rácsmintában helyeztük el (4.3. ábra), a többi kép esetében az objektumok rendezetlenül helyezkednek el a képtérben.

A tesztben szereplő randompont sztereogrammok sztereó módon megjelenítve egy-egy geometriai objektumot tartalmaznak a kép közepén. Mivel a randompont



4.3. ábra. Példa a második teszt mesterségesen generált képeiből (jobb szem képe, mélységtérképpel)

sztereogrammon monó megjelenítés esetén felismerhető formát nem találhatunk monó képmegjelenítés esetén e képek objektumon kívüli része fekete homogén színezést kapott. A tesztben szereplő randompont sztereogrammok sztereó módon megjelenítve egy-egy geometriai objektumot tartalmaznak a kép közepén. Mivel a randompont sztereogrammon monó megjelenítés esetén felismerhető formát nem találhatunk monó képmegjelenítés esetén e képek objektumon kívüli része fekete homogén színezést kapott.

A második teszt során tehát 26 db képből álló képi adatbázis állt rendelkezésre. Mivel a mesterséges képek meglehetősen hasonlítanak egymásra, a témák változatossága nem kellő mértékben biztosítható, így a képeket a résztvevők mindig azonos sorrendben tekintették meg, hogy az esetlegesen emlékezet miatt keletkezett torzulások azonosak legyenek. A képeket két képszekvenciába rendeztük, melyeket a résztvevő vagy monó vagy sztereó módon tekintett meg. A második teszt képei közül a teszt egy résztvevője, az első teszthez hasonlóan, minden képet pontosan egyszer tekintett meg, a képek felét monó, másik felét sztereó módon megjelenítve.

4.6.2. Résztvevők

A második teszt során a résztvevőket, az első teszthez hasonlóan, két csoportba sorolhatjuk, attól függően hogy a képsorozat első („A” csoport) vagy második („B” csoport) felét látták sztereó megjelenítési módban. A tesztet összesen 6 jól látó, megfelelő sztereó látású résztvevő végezte el, akik közül 3-an az „A” és 3-an a „B”

csoportba tartoznak. A résztvevők között 2 nő volt. A tesztet („B” csoport) egy gyengébb sztereó látású résztvevővel is lefolytattuk.

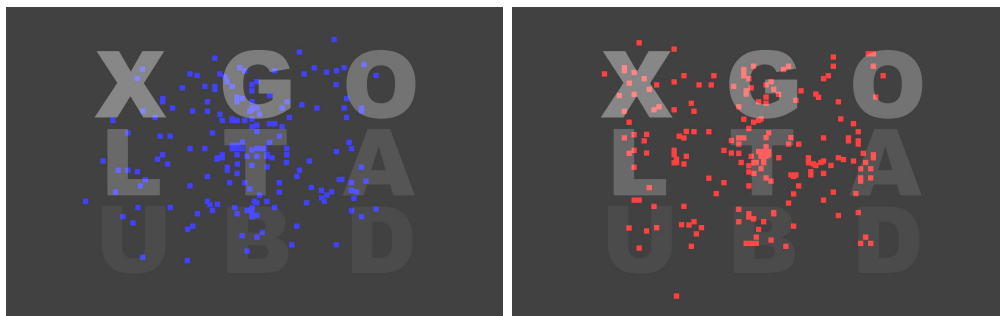
5. Adatelemzés, eredmények

A két vizuális teszt eredményeként rögzített adatok előzetes feldolgozása során összesen 47279 rekordos adatbázist építhettünk fel. Az előfeldolgozás során a kalibráció ellenőrző képeken mért elmozdulással korrigáltuk a rögzített fixációs pontok és szakkádok térbeli koordinátáit. Néhány rekordot pl. a képterületen kívül eső térbeli koordináták, vagy túl nagy korrekció szükségessége miatt (fej nagymértékű elmozdulása) érvénytelennek minősítettünk. Az előfeldolgozott adatbázis különböző típusú rekordjainak számát az 5.1. táblázat mutatja be.

5.1. táblázat. A tesztek eredményeiből felépült adatbázis rekordjainak száma

		fixációs pont		szakkád		összes rekord	
		összes	érvényes	összes	érvényes	összes	érvényes
első teszt	monó	8059	7697	8060	5728	16119	13425
	sztereó	7982	7700	8000	5230	15982	12930
	összesen	16041	15397	16060	10958	32101	26355
második teszt	monó	3860	3710	3863	2527	7723	6237
	sztereó	3726	3675	3729	2300	7455	5975
	összesen	7586	7385	7592	4827	15178	12212
mindösszesen		23627	22782	23652	15785	47279	38567

Az adatok előfeldolgozása során a rekordok térbeli koordinátáinak korrekciója, illetve helyességének (érvényességének ellenőrzése mellett) az egyes rekordhoz különböző statisztikai adatokat, illetve a térbeli koordináták alapján térbeli jellemzőket társítottunk. Az adatfeldolgozás első szintje után létrejövő adathalmaz rekordjai tehát egy résztvevő egy adott megjelenítési módban rögzített egy adott képéhez rendelik a fixációkat illetve a szakkádokat, megadva azok térbeli és idő-



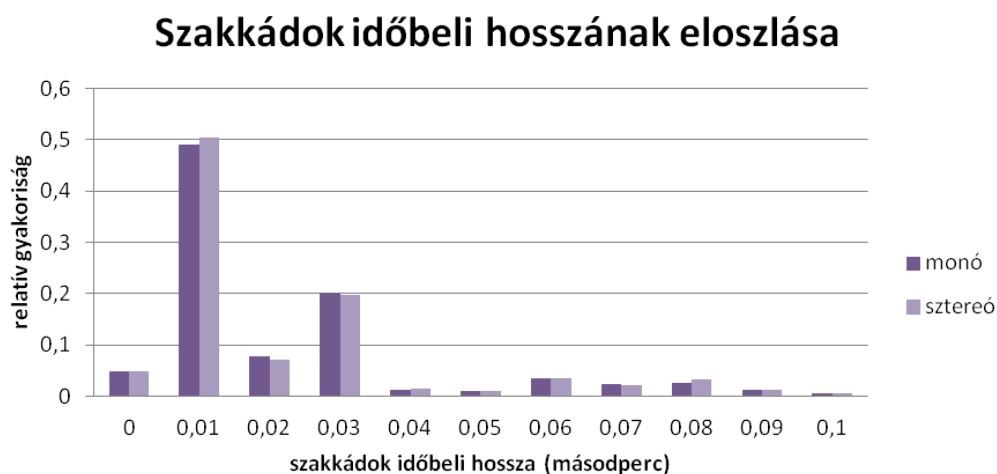
5.1. ábra. A második teszt egy mesterséges képének monó (balra) illetve sztereó fixációs pontjai

beli elhelyezkedését, illetve az éppen megtekintett képterület egyes jellemzőit (pl. színtelítettség, lokális kontraszt, mélységi határvonaltól, élektől való távolság). Az adatbázis adatszerkezete tehát lehetővé teszi, hogy a rekordokat résztvevő, kép, megjelenítési mód, idő szerint leválogassuk illetve szűrjük, vagy statisztikai elemzésnek vessük alá.

5.1. Szakkádok időbeli hosszának vizsgálata

A szakkádok a szem két fixáció közötti gyors mozgásai, melyek során az egyes újabb megtekinteni kívánt területek fókuszba, a látómező közepére kerülnek. A szakkádok során információ felvétel nem történik, így kutatásunk során a szakkádok komolyabb tanulmányozásával nem foglalkoztunk, az adatminták közül a fixációs pontokra koncentráltunk. Egy kérdés azonban mindenképp felmerülhet a szakkádokkal kapcsolatban. Az agy sztereoszkópikus képmegjelenítés esetén a monó megjelenítési módhoz képest nehezebben állítja elő az észlelt képet (a két kép összeecsúsztatása, regisztrációja szükséges) – kérdéses lehet, hogy a többlet információ feldolgozásának szükségessége hatással van-e a szakkádok időtartamára.

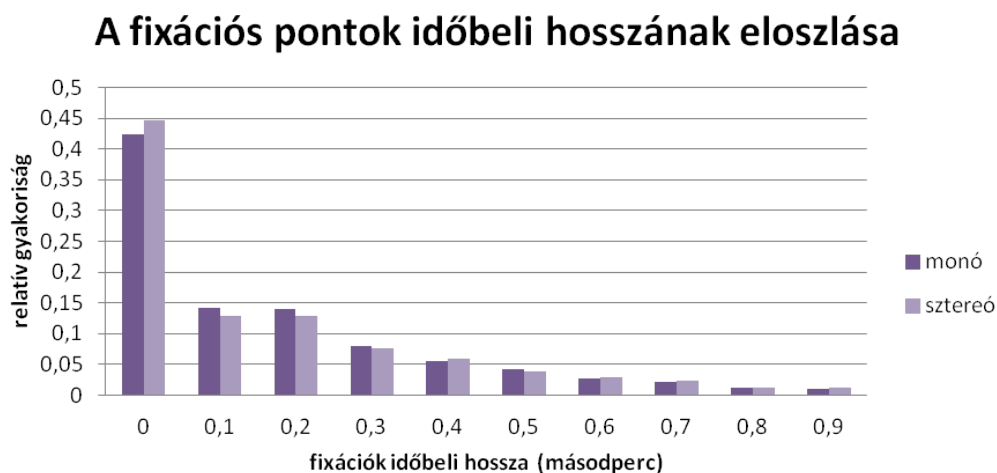
Kísérleteink tapasztalatai azt mutatják (5.2. ábra), hogy a képmegjelenítési mód változása nincs jelentős hatással a szakkádok időbeli hosszára. A grafikonról leolvasható, hogy a szakkádok hosszának módusza 0.01-0.02 másodperc közötti tartományra esik, mely közelítőleg megegyezik az irodalmi példákban leírtakkal [9].



5.2. ábra. A szakkádok időbeli hosszának eloszlása a rövid időtartományokon

5.2. Fixációs pontok időbeli hosszának vizsgálata

Hasonló módon vizsgálhatjuk a fixációk időbeli eloszlásának esetleges változását is. A fixációk a szem megnyugvásai egyes képterületeken, általában a szakkádoknál nagyságrenddel hosszabb ideig tartanak. Tesztjeink során rögzített fixációk időbeli hosszának eloszlását az 5.3. ábra szemlélteti.



5.3. ábra. A fixációk időbeli hosszának eloszlása a rövid időtartományokon

Az ábráról is leolvasható, hogy a képmegjelenítési mód változása a fixációk időbeli hosszát sem változtatja. Azaz a sztereó képek nehezebb feldolgozása a szem alapvető mozgásaira nincs közvetlen hatással.

5.3. Mélységi határvonal távolság vizsgálat

A fixációs pontok alapvető időbeli tulajdonságainak vizsgálata után a fixációkhoz tartozó képi jellemzők vizsgálata szükséges. Hipotézis: a fixációs pontok mélység észlelés (sztereó képmegjelenítés) esetében távolabb esnek a mélységi határvonalaktól, mint monó megjelenítés esetén.

A hipotézis vizsgálatához szükséges a tesztképekhez tartozó mélységtérképek ismerete. Az első teszt során nem álltak rendelkezésre diszparitás információk, így a vizsgálatot a képen való erős élek távolságával folytattuk le (a markáns élek általában jól korrelálnak a mélységi határvonalakhoz). A vizsgálatok során a fixációs pontok élektől való távolságának eloszlásában lényeges különbséget nem találtunk a monó és sztereó adatok között [7].

A második teszt képeinél rendelkezésre álltak diszparitás (mélységi) információk, így a továbbiakban a második teszt eredményeire szorítkozunk. A mélységtérképeken Canny él detektáló operátor használatával kerestük meg a markáns mélységi határvonalakat. Ezután egy úgynevezett távolság transzformáció, *distance transzformáció* segítségével a kép adott pontjaihoz meg tudtuk határozni a legközelebbi mélységi határvonaltól számított távolságot. Mivel a fixációs pontok helyzete a tesztek során kb. 30 pixeles pontatlansággal határozható meg, így a fixációs pontok helyzetét 30×30 -as négyzetben valószínűsítjük, az ezen belül található képjellemző értékeket átlagoljuk (5.4. ábra).

A mélységi határvonalaktól való eltérés vizsgálatánál tendenciális különbséget találtunk a teszt első felét sztereóban, illetve monóban megtekintő résztvevő csoport között. A különbség okainak feltárása jelenleg is folyamatban van, ezt okozhatták pl. a teszt szegényes témavilágú képei, vagy a teszt végére kialakuló fáradtság, de akár a sztereó megjelenítés okozta mélységérzet megjelenése, illetve „elvesztése” is. Az okok pontosabb vizsgálatához kísérlet megismétlését tervezzük, azonos képekkel és résztvevőkkel, a csoportkialakítások megfordítása mellett.



5.4. ábra. Egy tesztkép, annak mélységi határvonalai és a distance transzformáció eredménye, a fixációs pontok helyének jelölésével.

5.4. Bejárt képterület vizsgálat

A fixációs pontok eloszlásának vizsgálatánál fontos szempont lehet a bejárt képterület vizsgálata. Hipotézisünk szerint sztereó megjelenítés esetén a fixációs pontok klaszterezettebben, kisebb csoportokba rendeződve helyezkednek el egy képen, ezáltal a bejárt képterület csökkenhet. A hipotézisnek előzetes adataink alapján (első teszt) mind résztvevőnként, mind képenként lehet létjogosultsága – résztvevőnkénti feldolgozáson egy résztvevő adott képre vonatkozó fixációs pontjainak elhelyezkedését (vagyis a fizikailag a szem által bejárt képterületeket), képenkénti feldolgozáson az egy képre vonatkozó összes fixációs pontot értve.

A bejárt képterületeken a fixációs pontok körüli meghatározott sugarú területeket értjük (75 - 150 pixel sugarú környezeteket alkalmaztunk), a képhez tartozó fixációs pontok területeinek összesítése után képet kaphatunk a kép bejárásának mértékéről. Egy tesztképet, és annak egy résztvevő általi bejárását szemlélteti az 5.5. ábra. A vizsgálatnak a dolgozat benyújtásának idejéig számszerűen kife-



5.5. ábra. A második teszt egy képe, és annak egy résztvevő bejárása által generált maszkja

jezhető eredménye nincs, a sztereó adatok e tulajdonsága a képtartalom függvényében további elemzésre szorul.

5.5. Diszparitás szinteken töltött idő eloszlásának vizsgálata

A fixációs pontok eloszlását vizsgáltuk mélységi szintektől való függés szempontjából is. A vizsgálatokat a mélységi információ szükségessége miatt az előzőekhez hasonlóan csak a második teszt képein tudtuk elvégezni. Természetesen monó megjelenítés esetén a mélységi szinteket a résztvevők a teszt során nem láthatták (a tesztképek nagy részén sehogyan sem észlelhatték), így hatása önmagában nem lehetett befolyással a monó fixációs pontokra. Kérdés, sztereó megjelenítéskor a mélységi észlelés többlet információi befolyásolják-e az egyes mélységi szinteken töltött fixációs időt.

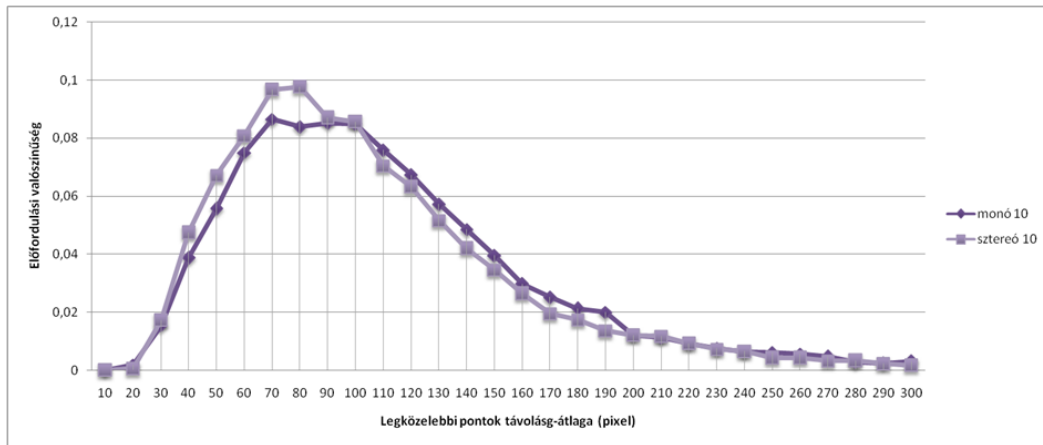
Érdekes kérdés lehet, hogy a diszparitás szintek közötti váltások (mélység ugrások) száma, hogyan változik a képmegjelenítési forma változásával. A diszparitás szinten töltött idők, illetve a diszparitás ugrások eredményei meglehetősen változatos eredményeket mutatnak. A továbbiakban az eredmények pontosabb elemzése után a változatos képek helyett szeparált csoportokat érdemes kialakítani (pl. diszparitás szintek szerint), és egy csoportból több képpel új kísérletet összeállítani, az esetlegesen a képtartalomtól (pl. diszparitás szintektől) függő tulajdonságok megállapítására.

5.6. Egymáshoz legközelebbi fixációs pontok távolsága

A fentiekben említett képjellemzőkön alapuló vizsgálatok bár mindenképpen a kutatás folytatásának további irányát jelentik, általában erősen képtartalom-függő eredményeket generáltak. Célszerű tehát a lehető legegyszerűbb adatszerkezetből a fixációs pont térbeli koordinátáiból kiindulni. Hipotézisünket, miszerint a fixációs pontok sztereó megjelenítés esetén klaszterezettebben helyezkednek el, egyrészt a következőképpen egyszerűsíthetjük – sztereó képmegjelenítés esetén a legközelebbi néhány fixációs pont távolsága kisebb, mint monó megjelenítés esetében.

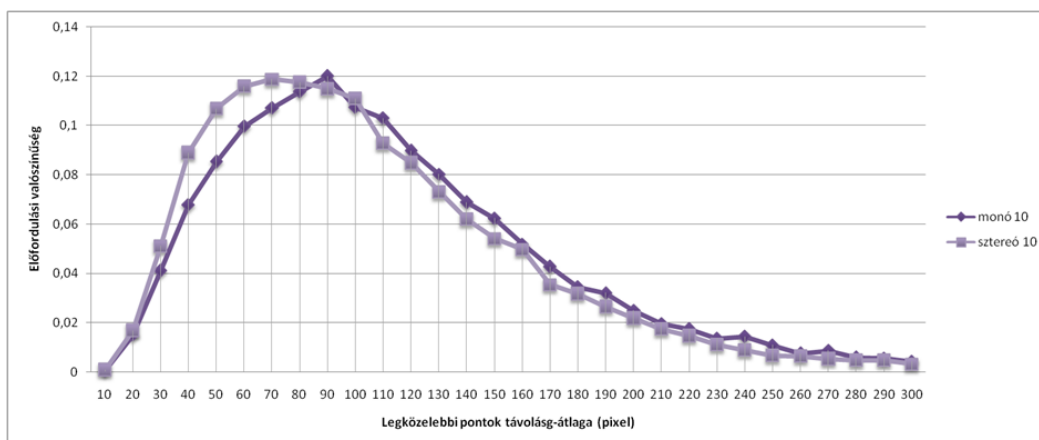
A vizsgálatnál a fixációs pontokat $f(x, y)$ vektorként definiáljuk, ahol x, y a vektor koordinátái a képtérben. Ebben az esetben az $f_j(x_j, y_j)$ j . fixációs pontra a vizsgálatnál alkalmazott módszert az 5.1. képlet írja le, ahol F az adott képre vonatkozó fixációs pontok halmaza.

$$tavolsag(f_j) = \min_{i \in I} \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{|I|}; f_i, f_j \in F; i \neq j; |I| = 0.1|F| \quad (5.1)$$



5.6. ábra. A legközelebbi fixációs pontok távolságának eloszlása

A vizsgálat eredményeként az 5.6. ábrán látható grafikont kaphatjuk. A grafikonon az egyes távolság értékek előfordulási valószínűségeit ábrázoltuk. Látható,

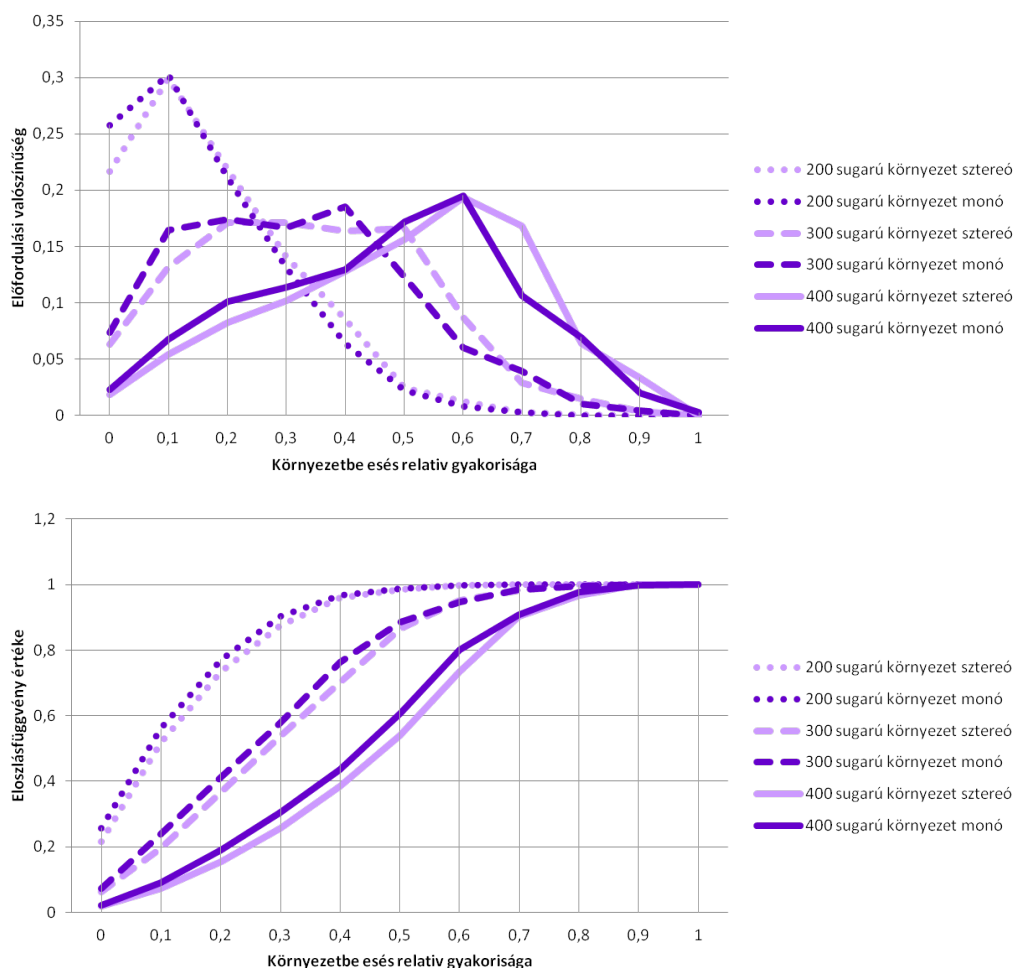


5.7. ábra. A legközelebbi fixációs pontok távolságának eloszlása, résztvevőnként kiértékelve

hogy a monó értékek esetében a grafikon csúcsa kis mértékben jobbra tolódik, vagyis a monó adatok esetében a legközelebbi fixációs pontok távolsága kis mértékben nagyobb a sztereó adatokhoz képest.

A kiértékelést résztvevőnként is elvégezhetjük, vagyis jelen esetben azt vizsgáljuk, a kép egy megtekintése, bejárása során változnak-e a legközelebbi fixációs pontok távolságai, vagyis az 5.1 képlet annyiban módosul, hogy F egy kép egy megtekintéséhez tartozó fixációs pontjainak halmazaként értelmezhető. Az eredmények ez esetben az 5.6. ábrán látható módon alakulnak. Az grafikon az előzőhöz hasonló jellegű, vagyis elmondható, hogy van különbség a fixációs pontok távolságában monó, illetve sztereó megjelenítés esetén.

A két függvény különbségének számszerű leírására a ferdeség (*skewness*) statisztikai leíró operátort választottuk, mely egy eloszlás szimmetriáját vizsgálja [26]. A vizsgálat eredményéül kapott eloszlások esetében a sztereó eloszlást mindig nagyobb pozitív ferdeség érték jellemzi mint a megfelelő monó eloszlást – azaz a sztereó görbék kissé balra tolódottabbak, a legközelebbi fixációs pontok sztereó esetben kisebb távolságra helyezkednek el, mind képenkénti kiértékelés, mind kizárólag a résztvevők saját adatainak figyelembe vétele mellett.



5.8. ábra. A 200,300,400 sugarú környezetbe eső fixációk számának eloszlása, illetve eloszlásfüggvénye

5.7. Fix sugarú környezetbe eső fixációs pontok

Amennyiben a fixációs pontok koncentráltóságát, kisebb csoportokba rendeződését szeretnénk vizsgálni, másik lehetőségként az egy fixációs pont körüli r sugarú körben elhelyezkedő az adott képre vonatkozó további fixációs pontok számát vizsgálhatjuk (5.2. képlet). A képlet eredményeként minden fixációs ponthoz meghatározzuk, hogy az adott kép pontjainak mekkora hányada esik a pont r sugarú környezetébe, ez alapján a fixációs pontokat tíz pontarány kategóriába soroljuk. A kiértékelést az előző pontban bemutatott módszerhez hasonlóan képenként, illetve

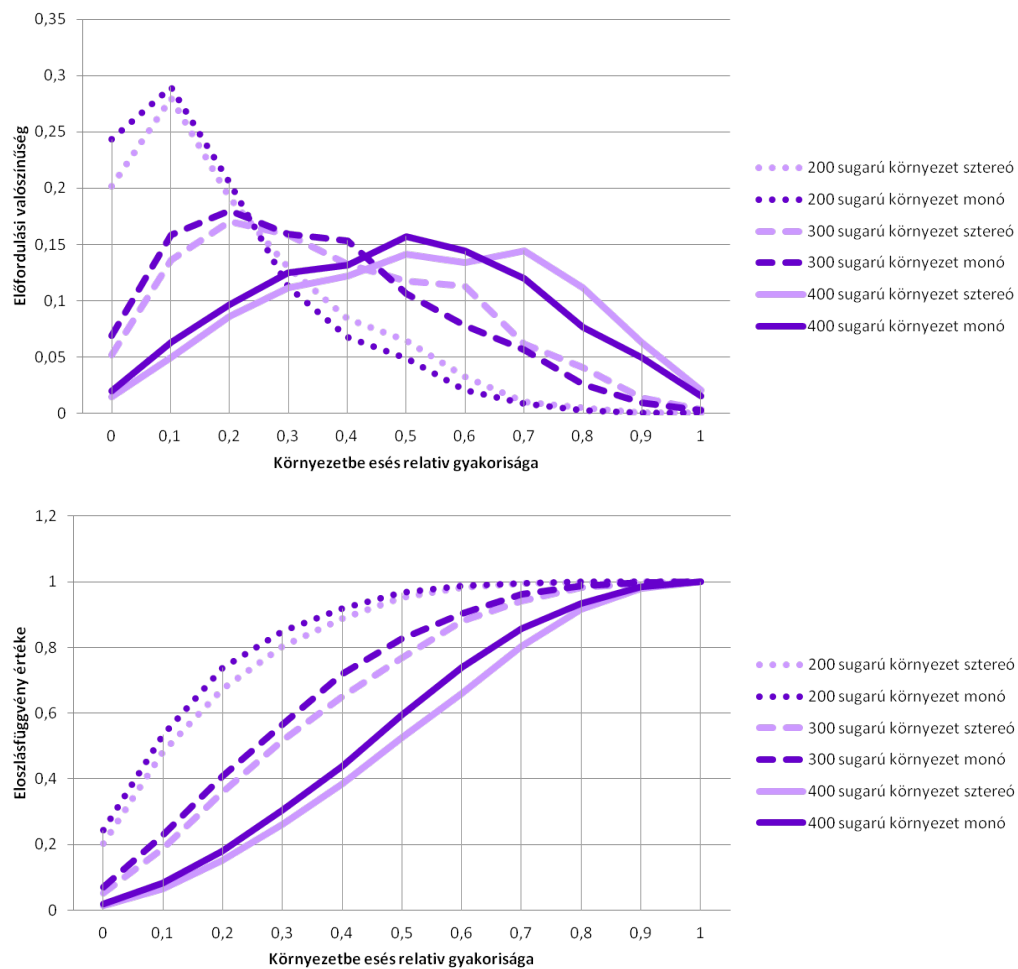
csak a résztvevők saját adatainak figyelembe vételével is lefolytathatjuk, a különbség ismét az F alaphalmaz értelmezésében adódik. A vizsgálat során képenként az egyes pontarány kategóriákba eső pontok relatív gyakoriságát számítottuk és ábrázoltuk.

$$pontokaranya(f_j) = \frac{|A|}{|F|}; A \subseteq F; f_i \in A | \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq r \quad (5.2)$$

A képenkénti kiértékelés eredményét az 5.8. ábra szemlélteti, ahol az r paraméter értékét 200,300, ill. 400 pixel értékben határoztuk meg. A grafikonról leolvasható, hogy a különböző pontarány kategóriákba (az adott kép fixációs pontjainak mekkora része esett bele az r sugarú környezetbe) eső fixációs pontok relatív gyakoriságánál a sztereó adatok kis mértékben a monó görbékhez képest jobbra tolódva helyezkednek el, azaz sztereó képmegjelenítés esetén átlagosan több fixációs pont kerül egy fixáció azonos sugarú környezetébe. A tendenciális eltérés az adatok eloszlásfüggvényének vizsgálatán keresztül még jobban szemléltethető, ahol látható, hogy a sztereó görbe folyamatosan a monó görbe alatt halad.

Az adatok kiértékelését a távolság alapú vizsgálatához hasonlóan résztvevőnként is elvégeztük. Ekkor tehát F halmaz az adott képhez tartozó egy résztvevő által generált fixációs pontok halmazának értelmezhető, az így kapott eredményeket az 5.9. ábra szemlélteti. Látható, hogy az eredmény hasonló a képenkénti kiértékeléshez – vagyis nem csak a képen található összes fixációs pontban, de egy kép bejárásánál is a sztereó adatok csoportba rendeződötten helyezkednek el a monó pontokhoz képest.

A különbségeket az első és a második teszt képeire egyaránt tendenciálisan kimutathatjuk, azonban az eltérés számértékben meglehetősen kismértékű, matematikai szignifikancia operátorokkal nehezen értelmezhető.



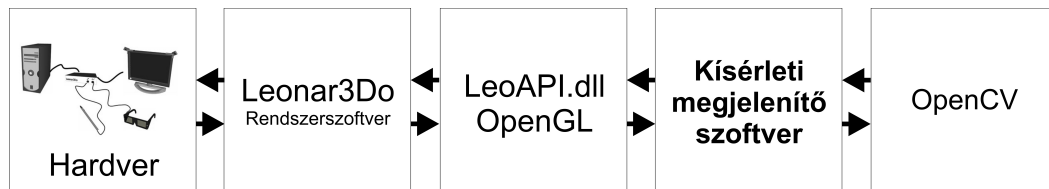
5.9. ábra. A 200,300,400 sugarú környezetbe eső fixációk számának eloszlása, illetve eloszlásfüggvénye résztvevőnkénti kiértékelés esetén (csak a saját pontok figyelembe vételével)

6. Szoftvertechnikai megvalósítás

A feladathoz kapcsolódóan a kísérleteket, illetve a rögzített adatok kiértékelését kellett szoftveresen támogatni, így három szoftverkomponens került megtervezésre és implementálásra. A vizuális kísérletekhez sztereó megjelenítést is támogató megjelenítő szoftvert implementáltam a Leonar3Do rendszerhez, valamint szoftverkomponenseket igényelt a második tesztnél felhasznált mesterségesen generált képek elállítása, illetve a tesztek során rögzített adatok kiértékelése is.

6.1. Sztereó megjelenítő szoftver vizuális tesztekhez

A kísérleti képmegjelenítés során aktív folyadékkristályos sztereó megjelenítést, a Leonar3Do rendszert alkalmaztuk. A képmegjelenítő szoftvert tehát ehhez a rendszerhez kellett illeszteni. A szoftvernek előre meghatározott képszekvenciákat kell időzítve lejátszani oly módon, hogy a monó illetve sztereó megjelenítési mód képenként beállítható legyen. A program C++ nyelven Visual Studio 2010 környezetben készült a következőkben bemutatott programkomponensek felhasználásával – a program alapvető szerkezetét a 6.1. ábra szemlélteti.



6.1. ábra. A sztereó képmegjelenítő program blokkvázlata

6.1.1. Leonar3Do

A Leonar3do a sztereó képmegjelenítő rendszerben alapvető hardverelemként vesz részt, a számítógép a szemüveg és a monitor szinkronizációját biztosítja, szolgáltatja a szemüveg elhelyezkedésének adatait, valamint biztosítja a sztereó és a monó megjelenítési formák közötti váltás lehetőségét. A Leonar3Do hardverrel a csomaghoz mellékelte LeoAPI (dll fájl) függvényeinek segítségével lehet kommunikálni. Az esetleges teszt közbeni monó-sztereó váltások miatt a szemüveget mindkét megjelenítési mód esetén viselték a résztvevők, monó esetben kikapcsolt (mindkét szem számára folyamatosan áteresztő) állapotban.

6.1.2. OpenGL

Az (*Open Graphics Library*) egy széles körben elterjedt a grafikai szabvány, amelyet a Silicon Graphics dolgozott ki. Fő alkalmazási területe a számítógéppel segített mérnöki-, gépészeti- tervezési eljárások, illetve a virtuális valóság alkalmazások. Számos platformon szimulátorok és videójátékok esetében is használják. Ez egy olyan API-nak tekinthető, amely biztosítja, hogy egy grafikus kártyát egyszerű felületen lehessen vezérelni, például háromdimenziós objektumokat kirajzolni és mozgatni. Használatához magas programozói felkészültség szükséges, azaz alacsony szintű API-nak tekinthető. A főbb lehetőségei: 3D-s objektumok létrehozása, elforgatása, eltolása, textúrázása, színezése, a fények kezelése [25].

Célszerű felhasználni a sztereó képmegjelenítő rendszerekben, mivel a Leo hardver segítségével képes a LeoAPI OpenGL objektumok kirajzolására, megjelenítésére.

6.1.3. OpenCV

Az (*Open Source Computer Vision Library*) egy képfeldolgozást és megjelenítést támogató függvény csomag, amelyet az Intel fejlesztett ki. A csomag előre implementált függvényeket tartalmaz a sokféle képfeldolgozási műveletre (például egyenes illesztés, elmosás, de még arcfelismerés is), s e függvényeket emellett a processzorra optimalizálták. A csomag függvényei között találhatók sztereó képalkotással foglalkozó függvények, ezek felhasználását azonban akadályozta, hogy

nagy részük számít a (sztereó képpár felvételekor rögzített) kamera kalibrációs adatokra. A tesztképeknél nem rendelkezünk ilyen kalibrációs adatokkal, ezért ezeket a függvényeket nem lehetett alkalmazni. Jól használhatók voltak viszont az OpenCV előre implementált függvényei a képmanipulációs műveleteknél, a képek betöltésénél, mentésénél, így az általam készített sztereó képmegjelenítő rendszer az OpenCV által használt *IplImage* formátumban tárolja a betöltött képeket.

A sztereó diavetítést biztosító programot OpenGL és OpenCV segítségével alapon valósítottam meg C++ nyelven, a sztereó megjelenítést a LeoAPI-n keresztül a Leonar3Do végzi. A program az előre definiált .txt kiterjesztésű szövegfájlban meghatározott képeket szekvenciális sorrendben jeleníti meg. A képekhez a fent említett fájlban egyedileg definiálható a megjelenítési idő, és a beolvasás illetve a megjelenítés módja.

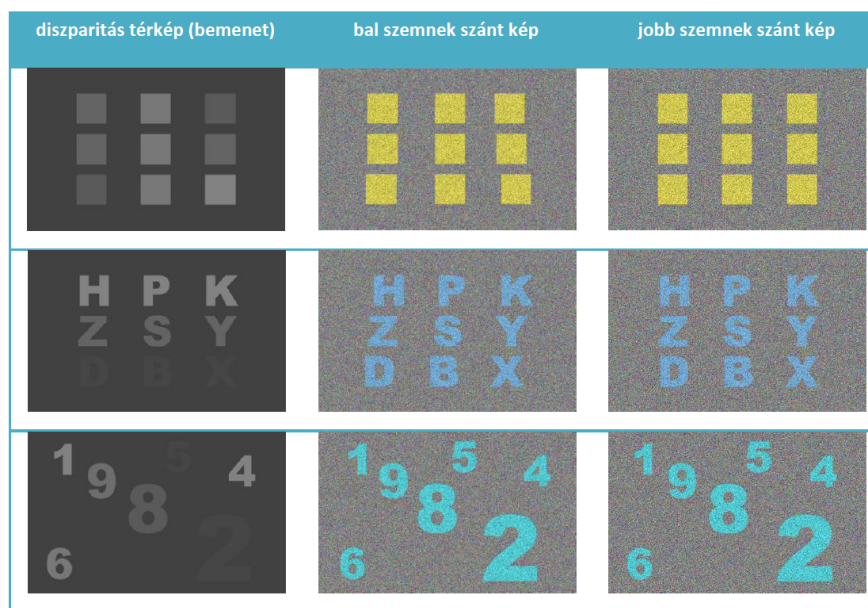
A program használatakor az adott képernyő felbontását meg kell adni, esetünkben ezek a felhasznált Samsung monitor legnagyobb felbontása (1680×1050 pixel). A képek a képernyőfelbontás helyes beállítása után eredeti méretben (a kép egy pixele a monitoron egy pixelként) jelennek meg a képernyő bal felső sarkához illesztve.

A program sztereó képpárok beolvasására képes, azonban minden képhez megadható olyan paraméter mely szerint a képnek nincsen másik szemnek szánt párja, ebben az esetben az adott kép megjelenítéséhez csak egy kép-fájl beolvasása történik meg. A megjelenítési mód kiválasztása azonban független az egy képhez beolvasott fájlok számától – teljes beolvasott képpár (jobb és bal szemnek szánt kép fájl) esetén monó megjelenítésnél a jobb szemnek szánt képet láthatjuk, egy beolvasott kép-fájl esetén sztereó megjelenítésnél mindkét szem számára azonos kép jelenik meg. A program elsősorban a kutatás támogatására készült, azt a kísérletek operátora mint egyetlen azonosított aktor használja fel, így a program grafikus felület nélkül parancssorból indítható változatban készült el.

6.2. Képgenerálás és adatfeldolgozás szoftvertchnológiai vonatkozásai

6.2.1. Mesterséges képek generálása

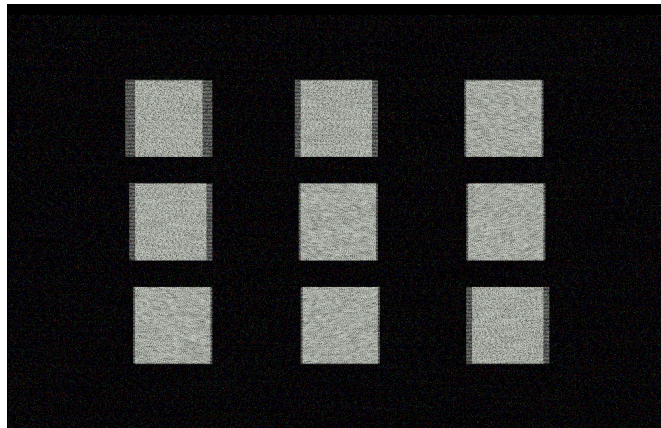
A második vizuális teszt kísérleteihez felhasznált mesterséges képek generálásához is megfelelő szoftverre volt szükség. Grafikai programok meglehetősen sokféle palettát kínálnak fel képek generálásához, azonban a szükséges random-pont textúrájú képek nagy mennyiségben történő előállítását saját szoftverrel végeztem. A szoftver bemenete a képek diszparitás térképe, a 0 diszparitást 100-as intenzitás értékként reprezentálva. A diszparitás térképen 100-nál kisebb intenzitás értékű objektumoknak a képenyő síkja mögött, a 100-nál magasabb intenzitás értékű objektumoknak az előtt kell a tesztek során megjelenni. A kísérletnél a homogén felületek helyett a véletlen textúrák alkalmazása segíti az agyban a képek regisztrációját ezáltal a térélmény kialakulását.



6.2. ábra. A kísérleti képgeneráló szoftver által előállított képek

Az OpenCV beépített függvényeit felhasználó C++ nyelvű szoftver a bementként megadott diszparitás térkép listához generálja a sztereó képpráok jobb illetve bal szemnek szánt megjeleníthető változatait. Az összehasonlíthatóság illetve a

rögzített adatok kiértékelésének megkönnyítése érdekében a jobb szem képén az objektumok helyzete megfelel a diszparitás térképnek, a megfelelő mértékű diszparitásokat (eltolásokat) a bal szemnek szánt képen alkalmaztam. Az objektumokat a kísérletek monotonitásának csökkentése érdekében különböző színekben jelenítettem meg – a színek, illetve az objektum-textúrák előállításánál fontos szempont volt, hogy azok ne homogén felületek legyenek, illetve megfelelően elváljanak a háttér textúrájától. Egy példabemenetre adott kimenetet a 6.2. ábrán mutatok be. A mesterségesen generált képek a korábban bemutatott sztereó képmegjelenítő rendszerrel a 6.3. ábrán látható módon jeleníthetők meg.



6.3. ábra. A mesterségesen generált kép a kísérleti képmegjelenítő szoftverrel megjelenítve – jól láthatóak az objektumok különböző diszparitás értékei. (A kép háttér textúrája a jobb szemléltetés érdekében sötétítve.)

A kísérletekben felhasznált random-pont sztereogrammok a szoftver módosított változatával készültek – ez esetben a háttér textúra egy darabja lett eltolva a bal szem képén a diszparitás térképnek megfelelően.

6.2.2. A kísérletek során rögzített adatok feldolgozásának támogatása

A kísérletek kiértékeléséhez a Képfeldolgozás Kuattólaboratóriumban korábban készült keretrendszer [7], mely képes a vizuális kísérletek által rögzített adatok betolvasására és összehasonlítások futtatására (a keretrendszer készítésénél részt vettem a szoftver által felépített adatszerkezet kialakításában). A dolgozatban be-

mutatott, általam kidolgozott összehasonlító algoritmusokat a keretrendszer adat-szerkezetének megfelelően C++ nyelven implementáltam, és a keretrendszerhez illesztettem.

A képgeneráló, képmegjelenítő és adatfeldolgozó szoftverek és szoftver-komponensek segítségével a kutatás során szükséges kísérletek elvégezhetőek voltak, további szoftverek elkészítésére a kutatás során nem volt szükség.

7. Összefoglalás

A képminőség objektív méréséhez olyan módszerre van szükség, mely a lehető legjobban közelíti az emberi látórendszer tulajdonságait. Az objektív módszerek közelítését a látórendszer sajátosságaihoz a következő adatok megállapításával lehet finomítani: tudnunk kell hova nézünk a képen (mik a fontos képterületek), ott mennyire vagyunk érzékenyek a torzulásokra, lokálisan mekkora a torzulás mértéke.

A jelenleg két dimenziós képekhez használt objektív képminőség mérési jelzőszámok akkor alkalmazhatóak változtatás nélkül sztereó megjelenítés esetére is, ha a képmegjelenítési mód változása a fontos (fókuszba kerülő) képterületeken, illetve az érzékenység mértékén nem változtat. Az torzulásokra való érzékenység vizsgálata több kutatás tárgyát is képezte [9, 30], azonban a fókuszba kerülő területek, azaz fixációs pontok eloszlásának általános vizsgálatára irodalmi adatokat nem találtunk. Célunk különbségek keresése monó és sztereó képmegjelenítési mód esetén a fixációs pontok eloszlásában.

A dolgozatban bemutatott kutatás során a monó és sztereó észlelés különbségeinek vizsgálatára két vizuális tesztet állítottunk össze, mely során a résztvevők szemmozgását figyelve következtethetünk az ember számára fontos területekre. A teszt során a képek szabad bejárását, azaz feladat nélküli megtekintését vizsgáltuk, majd a kapott mintákon statisztikai elemzést végeztünk. A tesztekbe 16 résztvevőt vontunk be. Az összesen közel 9 órán keresztül lezajlott kísérletek során összesen 7 GB-nyi adatot (képfelvétel, illetve fixációs pontok sorozata) rögzítettünk, melyek feldolgozásával egy 47279 rekordos adatbázist építettünk fel.

A tesztek eredményei alapján megállapítottuk, hogy nincs lényegi különbség a képet bejáró szemmozgást jellemző fixációs pontok és szakkádok időbeli hosszában. Ugyanakkor megállapítható, hogy a fixációs pontok elhelyezkedésében talál-

ható különbség a monó és sztereó látás mód adatai között. Sztereó megjelenítés esetén a fixációs pontok kis mértékben koncentráltabban, egymáshoz közelebb helyezkednek el a képen. Az észlelhető különbséget két megközelítésben vizsgáltuk, egyrészt a fixációs pontok fix sugarú környezetébe eső pontok számának vizsgálatával, ahol tendenciális eltérést tapasztaltunk, másrészt a legközelebbi fixációs pontok távolságának vizsgálatával, ahol a függvények ferdeségében számszerű eltérést azonosítottunk.

A különbség okainak feltárása céljából a rögzített adatokon vizsgáltuk a fixációk élektől, a mélységi határvonalaktól való távolságát, az egyes diszparitás szinteken eltöltött időt, a diszparitás ugrásokat egyaránt. Az adatok közötti különbségeket egyelőre meglehetősen képtartalom függőnek találtuk, a különbségek okainak pontos feltárására további vizsgálatok szükségesek.

Kutatásunk alapkutatás, mely során alapvető különbségeket kerestünk a monó és sztereó képek megtekintési módjaiban. A különbségek részletes feltárása után eredményeink alkalmazott kutatásban az objektív képminőség mérő pooling technikák súlytérképeinek sztereó esetben való kialakításánál, illetve a képkódolási technikák sztereó változatainak minőség érzet javításánál kaphatnak szerepet.

Irodalomjegyzék

- [1] Bakó Á.: Lassújel az Északi összekötőn. letöltve: 2011.05.08.
- [2] 3d display revenues forecast. http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xsl/100104_3d_display_revenues_forecast_to_reach_22_billion_by_2018.asp. megtekintve: 2012.04.18.
- [3] Tanács Attila – Czúni László: *Képi információ mérése*. 2011, Typotex Kiadó.
- [4] O. Chelnokova – B. Laeng: Three-dimensional information in face recognition: An eye-tracking study. *J Vis*, 11. évf. (2011) 13. sz.
- [5] Circular polarized 3d glasses. <http://www.circularpolarized3dglases.com/>. letöltve: 2012. február 8.
- [6] Ricardo L. de Queiroz – C. K. Choi – Young Huh – K. Raghunath Rao: Wavelet transforms in a jpeg-like image coder. *IEEE Trans. Circuits Syst. Video Techn.*, 7. évf. (1997) 2. sz., 419–424. o.
- [7] Barti G.: *A képi minőség mérése*. 2012, Szakdolgozat, Pannon Egyetem, Veszprém.
- [8] A Gotchev: 3d content and 3d media for future internet mobile devices. In *Japan EU Symposium on the New-Generation Network and Future Internet, 2009. September 9.* (konferenciaanyag). 2009.
- [9] Tien Ho-Phuoc – Nathalie Guyader – Frédéric Landragin – Anne Guérin-Dugué: When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision*, 12. évf. (2012. február) 2. sz.

- [10] Holografika. <http://www.holografika.com/>. megtekintve: 2012.04.18.
- [11] Chang Hong Liu – James Ward – Andrew W. Young: Transfer between two- and three-dimensional representations of faces. *Visual Cognition*, 13. évf. (2006) 1. sz., 51–64. o.
- [12] Kiss P. J.: *Sztereo képmegjelenítés minőségét befolyásoló tényezők*. 2011, Szakdolgozat, Pannon Egyetem, Veszprém.
- [13] Kiss P. J.: *Sztereo képek képminőségi jellemzőinek vizsgálata*. 2012, KTD dolgozat, Pannon Egyetem, Veszprém.
- [14] H. Jorke – A. Simon – M. Fritz: Advanced stereo projection using interference filters. In *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008* (konferenciaanyag). 2008. máj, 177 –180. o.
- [15] Luo Juan – Oubong Gwun: A comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing (IJIP)*, 2009., 143–152. o.
- [16] Bela Julesz: Foundations of cyclopean perception. Chicago: University of chicago press, 1971. *Behavioral Science*, 17. évf. (1972) 3. sz., 310–312. o.
- [17] Timor Kadir – Michael Brady: Saliency, Scale and Image Description. *International Journal of Computer Vision*, 45. évf. (2001. november) 2. sz., 83–105. o. ISSN 09205691.
- [18] E Charles Leek – Filipe Cristino – Lina I Conlan – Candy Patterson – Elly Rodriguez – Stephen J Johnston:.
- [19] Leonar3do csomag tartalma. <http://www.leonar3Do.com/>. megtekintve: 2011.05.23.
- [20] David G. Lowe: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60. évf. (2004) 2. sz., 91–110. o.
- [21] Látás kísérletek. <http://pszichologia.btk.pte.hu/letolt/03adaptacio.pdf>. letöltve: 2011. május 8.

- [22] Dragorad Milovanovic – Zoran Bojkovic: Trend and challenges in 3d digital television: video formats and transmission standards. In *Proceedings of the 14th WSEAS international conference on Communications*, ICCOM'10 konferenciasorozat. Stevens Point, Wisconsin, USA, 2010, World Scientific and Engineering Academy and Society (WSEAS), 127–134. o. ISBN 978-960-474-200-4. 8 o.
- [23] Anush K. Moorthy – Alan Conrad Bovik: Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 20. évf. (2011) 12. sz., 3350–3364. o.
- [24] Michael C Mozer – Adrian Fan: Top-down modulation of neural responses in visual perception: a computational exploration. *Natural Computing*, 7. évf. (2008) 1. sz., 45–55. o.
- [25] Opengl. <http://www.opengl.org/>. megtekintve: 2011.05.23.
- [26] Michelberger Pál – Szeidl László – Várlaki Péter: *Alkalmazott folyamatstatisztika és idősor-analízis*. 2001, Typotex Kft.
- [27] Majid Rabbani – Paul W. Jones: *Digital Image Compression Techniques*. 1991, SPIE – The International Society for Optical Engineering, Ballingham, Washington, USA.
- [28] Recommendation ITU-R BT.500-13. <http://www.itu.int/rec/R-REC-BT.500-13-201201-I/en>. letöltve: 2011.10.02.
- [29] Sekuler Robert – Blake Randolp: *Észlelés*. 2000, Osiris Kiadó.
- [30] Pieter Seuntiens – Lydia Meesters – Wijnand Ijsselsteijn: Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric jpeg coding and camera separation. *ACM Trans. Appl. Percept.*, 3. évf. (2006. április) 2. sz., 95–109. o. ISSN 1544-3558. 15 o.
- [31] Pieter Seuntiëns – Lydia Meesters – Wijnand Ijsselsteijn: Perceptual evaluation of JPEG coded stereoscopic images. In *Storage and Retrieval for Image and Video Databases* (konferenciaanyag). 2003.

- [32] Aljoscha Smolic (szerk.): *Introduction to Multiview Video Coding*. 2008. Jan, International organisation for standardization, ISO/IEC JTC 1/SC 29/WG 11N9580, Antalya, Turkey.
- [33] Sztereó képpárok. <http://3dmedia.com/gallery/index.php>. megtekintve: 2011.10.11.
- [34] Sztereó képpárok. <http://finepix.com/3d/en/download/index.html>. megtekintve: 2011.10.11.
- [35] Sztereó képpárok. http://en.wikipedia.org/wiki/Fujifilm_FinePix_Real_3D_W1. megtekintve: 2011.11.08.
- [36] Sztereó képpárok. <http://www.steves-digicams.com/camera-reviews/fujifilm/finepix-real-3d-w3/sample-photos-122.html>. megtekintve: 2011.09.23.
- [37] Sztereó képpárok. <http://www.neocamera.com/camera/fuji/w3/samples>. megtekintve: 2011.10.11.
- [38] Andrew S. Tanenbaum: *Számítógép hálózatok*. 2004, Panem Könyvkiadó Kft., Budapest.
- [39] E. Trucco – A. Vetri: *Introductory Techniques for 3-D Computer Vision*. 2008, Prentice Hall.
- [40] Anthony Vetro: Representation and coding formats for stereo and multiview video. In ChangWen Chen – Zhu Li – Shiguo Lian (szerk.): *Intelligent Multimedia Communication: Techniques and Applications*. Studies in Computational Intelligence sorozat, 280. köt. 2010, Springer Berlin Heidelberg, 51–73. o. ISBN 978-3-642-11685-8.
- [41] J. J. Vos: Some new aspects of color stereoscopy. *J. Opt. Soc. Am.*, 50. évf. (1960. Aug) 8. sz., 785–790. o.
- [42] Zhou Wang – A.C. Bovik – H.R. Sheikh – E.P. Simoncelli: Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13. évf. (2004. ápril) 4. sz., 600–612. o. ISSN 1057-7149.

- [43] Zhou Wang – Qiang Li: Information Content Weighting for Perceptual Image Quality Assessment. *IEEE Transactions on Image Processing*, 20. évf. (2011. május) 5. sz., 1185–1198. o. ISSN 1057-7149.
- [44] Zhou Wang – Xinli Shang: Spatial pooling strategies for perceptual image quality assessment. In *IEEE Int’l Conf. on Image Processing* (konferenciaanyag). 2006.
- [45] Zhou Wang – Eero P. Simoncelli – Alan C. Bovik: Multi-scale structural similarity for image quality assessment. In *in Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers, (Asilomar* (konferenciaanyag). 2003, 1398–1402. o.
- [46] L. P. Yaroslavsky: On redundancy of stereoscopic pictures. In *Acta Polytechnica Scandinavica, n. 149. Image Science’85. Proceedings. Helsinki, Finland., 11-14 June* (konferenciaanyag). 1985, 82–85. o.
- [47] L. P. Yaroslavsky – J. Campos – M. Espínola – I. Ideses: Redundancy of stereoscopic images: Experimental evaluation. *Opt. Express*, 13. évf. (2005. Dec) 26. sz., 10895–10907. o.

CD melléklet tartalomjegyzéke

- diplomaadolgozat
- internetes források
- kísérleti képmegjelenítő szoftver
- kísérleti képgeneráló szoftver
- kísérleti kiértékelő szoftver
- kísérleti képek és képszekvenciák
- kísérleti eredmény adatbázisok