

# Free Viewpoint Television: új perspektíva a 3D videó továbbításban

HUSZÁK ÁRPÁD

*Budapesti Műszaki és Gazdaságtudomány Egyetem*

*huszak@hit.bme.hu*

*Kulcsszavak: 3D videó, Free Viewpoint Video, képszintetizálás, streaming*

**A 3D multimédia folyamatos fejlődésen megy keresztül napjainkban, melynek legújabb irányzata a FVV (Free Viewpoint Video). A technológia lehetőséget ad a felhasználónak, hogy valós időben tetszőleges nézőpont alapján jelenítsék meg a videótartalmat. Tetszőleges nézőpontok előállítására képszintetizáló algoritmus segítségével történik, amely sokkamerás felvételeken alapszik. Az új szolgáltatás megvalósításához számos részfeladatot kell megoldani, a képrögzítés, tömörítés, átvitel, képszintézis és megjelenítés területén. A cikk ezt a folyamatot hivatott áttekinteni. Ha a folyamatlanc minden eleme hibátlanul és kimagasló módon képes működni, csupán akkor várhatjuk, hogy a tartalom előállítók is támogassák a technológiát, melytől most valóban nagy áttörést várunk a 3D videózás területén.**

## 1. Bevezetés

A mélységérzetet visszaadó képmegjelenítés hosszú múltra tekint vissza, hiszen az első 3D technikát, a szín-szűrőes, más néven anaglif szemüveges módszert már több mint 160 éve, 1853-ban dolgozta ki Wilhelm Rollmann. Azóta számos eljárás és technológia látott napvilágot, melyek népszerűsége erősen ingadozott az évtizedek folyamán. Az elmúlt néhány évben az otthoni 3D TV-k elterjedésének lehettünk tanúi, azonban a kezdeti, erős marketinggel fűtött lelkesedés mára alábbhagyott. Igazán nagy áttörés tehát most sem született. A szín-szűrős módszert felváltotta a polárszűrős és aktív-zárás megoldás, azonban a kényelmetlenséget jelentő szemüveg használata továbbra is nélkülözhetetlen. Az igazi, maradandó áttörés a 3D televíziózás területén akkor várható, ha sikerül szemüveg nélkül is mélységérzetet visszaadó képmegjelenítést megvalósítani. Hasonló folyamatokat láthattunk a mozi esetén is, azonban ott a szemüveg okozta kényelmetlenséget jobban tolerálják a nézők.

A 3D megjelenítés tökéletesítése folyamatos motívációt ad a kutatóknak és a fejlesztőmérnököknek. Ma már megjelentek olyan szemüveg nélküli is nézhető 3D kijelzők prototípusai, mely reményeink szerint már valóban meghozzák az áttörést. Az új technológia a sokkamerás felvételeken és a tetszés szerint nézőpont előállításán alapul.

A 3D videó új generációja a Free Viewpoint Video (FVV), melynek lényege, hogy felhasználók interaktív módon változtathatják a megjelenített nézőpontot, ahogyan például a 3D számítógépes grafika esetén már megszokhattuk. A legfőbb különbség azonban az, hogy míg az utóbbi esetén 3D modelleket alkalmazunk, a FVV esetén valós kamerák képét használjuk a képszintézishez. Akár apró nézőpont elmozdulásokat is kezelni kell a rendszernek, például a néző fejének mozgását követve. Rá-

adásul, a perspektíva változása folyamatos, így valószínűleg képszintézisre van szükség minden egyes néző számára. Az egyedi nézőpontok előállításához több nagyfelbontású kamera képe szükséges, melyek IP-hálózaton való továbbítása a FVV képszintetizáló egységhez újabb nehézséget jelenthet, főleg abban az esetben, ha maga a felhasználó eszköze végzi ezt a feladatot. Az előállított kép minősége annál jobb, minél közelebb helyezkednek el egymáshoz a kamerák, vagyis minél több kamerát alkalmazunk. Ugyanakkor, a nagyszámú kamerakép átviteléhez jelentős átviteli kapacitásra van szükség, ami hatékony átviteli megoldások nélkül nem lenne megvalósítható, hiszen a megnövekedett forgalom gyorsan túlterhelné a hálózatot. A hálózati kérdéseken túl természetesen a szemüveg nélküli kijelző is fontos szerepet játszik. Míg egyes megoldások perspektíva-függő 2D megjelenítést támogatnak, ma már megjelentek olyan szemüveg nélküli 3D kijelzők prototípusai, melyek valós mélységérzetet keltenek. Utóbbiban a budapesti székhelyű Holografika Kft. úttörő szerepet játszik.

A FVV iránti érdeklődés folyamatosan növekszik, hiszen ez a technológia már valóban meghozhatja az áttörést a 3D videózás területén. Számos újszerű alkalmazás megjelenését is várhatjuk az technológiától, ilyen például az IPTV alapokon működő Free Viewpoint Television (FTV). Az alkalmazott műszaki megoldás hasonló lehet a mai IPTV rendszerekhez, annyi különbséggel, hogy egy csatornához nem egy videó folyam tartozik, hanem akár több tíz is. A másik fontos különbség, hogy minden felhasználó egyedi perspektívából nézi a megjelenített tartalmat. Mivel a technológia még kutatási szinten van csupán, több egymással versengő megoldásról lehet olvasni, akár az átviteli megoldások, akár a megjelenítés terén.

Hasonlóan a hagyományos médiakommunikációs rendszerekhez, a kamerákkal rögzített videótartalom számos folyamaton halad keresztül, mire a felhasználó ké-

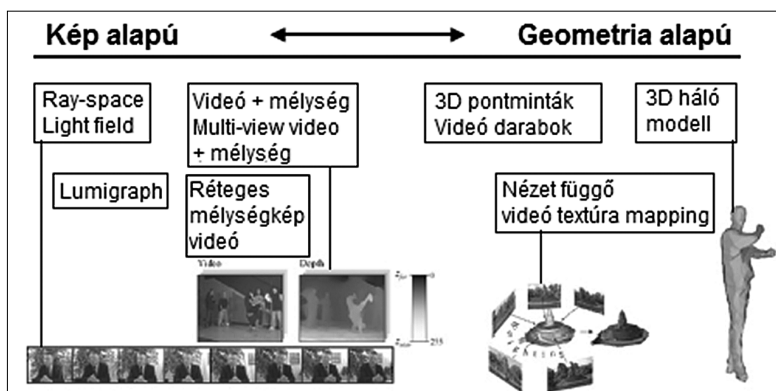
szülékén megjelenítésre kerül. A folyamat természetesen a tartalom előállításával, illetve rögzítésével kezdődik, majd a tömörítési eljárások segítségével az átvitelre és tárolásra alkalmas formátumba való alakítással folytatódik. A következő feladat a tömörített tartalom felhasználóhoz való továbbítása vezeték és vezeték nélküli hálózatokon, ezt követően pedig dekódolás, virtuális nézőpont szintézise, majd megjelenítés. Meg kell említeni, hogy a képsintézis (renderelés) funkciót több helyen is meg lehet valósítani, nemcsak a felhasználó eszközén, így például egy központi médiaszerveren vagy akár elosztott módon is.

A cikk következő szakaszaiban a FVV/FTV tartalom előállítás, tömörítés, átvitel és megjelenítés módszereit tekintjük át.

## 2. Többkamerás felvétel

A 3D tartalom előállításának több módja is létezik, mely alapulhat 3D hálómodelleken, kamerafelvételeken, illetve ezek kombinációján. A képalkotási módszereket ábrázoljuk akár egy egyenes mentén is, ahogyan az 1. ábra mutatja [1].

1. ábra FVV képalkotási módszerek



A geometria alapú 3D háló modellezési módszert már régóta alkalmazzák a számítógépes grafikában. A modellezés előnye, hogy a felépített virtuális világban a felhasználó tetszőleges nézőpontot választhat magának, bármiféle megkötés nélkül. A módszer hátránya, hogy nagy a számításigénye és a valós világ megfelelően részletes modellezése még várat magára.

A másik irányzat a Ray-Space vagy Light Field módszer [2], mely valós kameraképeket használ és a virtuális nézőpontban lévő képet a szomszédos képekből interpolációval állítja elő. A kamerák elhelyezkedésétől függően két vagy három kamera képére van szükség a tetszőleges nézőpont előállításához. A néző egy szűkebb térrészben választhatja ki a nézőpontját, azonban a szintetizált képek szinte megkülönböztethetetlenek lehetnek a valóságtól. A módszer hátránya, hogy nagyon sok kamerára van szükség, ami-

nek következtében hatalmas adatmennyiséget kell feldolgozni, illetve szükség esetén továbbítani.

A két szélsőséges megoldás között azonban vannak más lehetőségek is, ahogy azt az ábra is mutatja. A valós környezet rögzítésére használt másik módszer a hagyományos kamerák mellett mélységkamerákat is alkalmaz: MVD (Multiview Plus Depth) [3]. Az MVD megoldás előnye, hogy nagyobb távolságokra lehet elhelyezni a kamerákat, így kevesebb eszközre van szükség, mint a Light Field esetén. Mélységkamerák pedig ma már olcsón hozzáférhetőek, pl. Kinect, Wii.

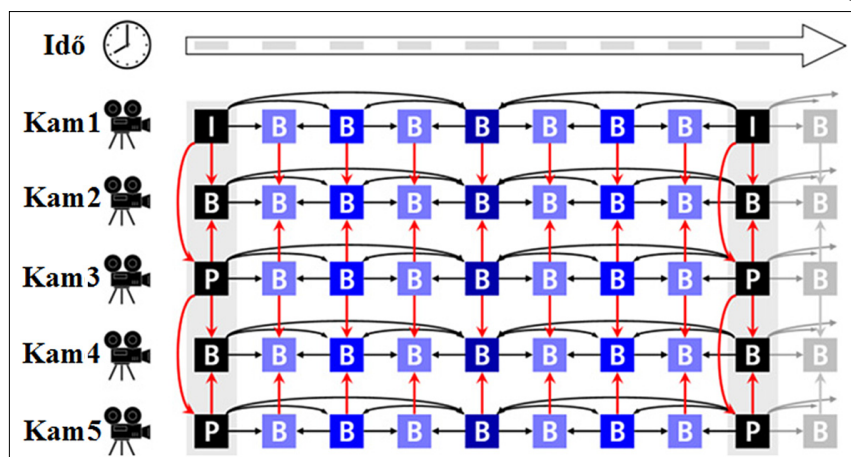
A többféle módszerhez különböző felvétel, feldolgozás és adatstruktúra tartozik. Mindegyiknek van előnye és hátránya, így minden egyes megvalósításnál meg kell fontolni, hogy éppen melyik megoldást érdemes használni a legjobb élmény érdekében.

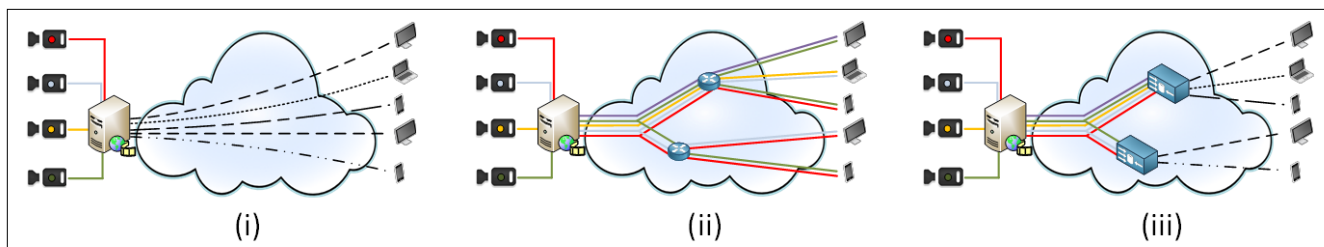
Természetesen a kamerák száma és elhelyezése megköti a virtuális perspektíva tartományát. A legegyszerűbb elrendezés, ha a kamerákat egy egyenes mentén (pl. sín) helyezük el, azonban a kamerák telepítése történhet síkon, rácspontos elrendezésben, vagy akár kupolaszerűen is.

## 3. Multiview Video Coding

A hálózati rendszerek korlátozott kapacitásai túl szűkösek ahhoz, hogy nagyszámú kamera- és mélységképet lehessen a hálózat túlterhelése nélkül a felhasználóknak továbbítani. Míg a hagyományos 2D videókódolási eljárások a képkockák közötti jelentős redundanciát használják ki, addig az új többkamerás videó tömörítő eljárások a térbeli hasonlóságon is alapszanak. A Multiview Video Coding (MVC) eljárás a különböző kameraképek között fellelhető jelentős hasonlóságot próbálja meg kihasználni azzal, hogy együtt kódolják a kamerák képeit. A módszer már a H.264/AVC szabvány része [4], amely jelenleg a leghatékonyabb tömörítési módja egy többkamerás videónak, de akár sztereoszkóp felvételek tömörítésére is kiválóan alkalmas. Az MVC működését a 2. ábra mutatja be.

2. ábra Multiview Video Coding





3. ábra FVV streaming modellek

Mélységképet és videóképet tartalmazó Free Viewpoint Video esetére az MPEG-C Part 3 [5] szabvány tartalmaz megoldást, azonban ezen a téren még folynak kutatások. A mérések alapján a mélységinformáció kb. 10-20%-kal növeli a szükséges sáv szélességet az átvitel során [6].

#### 4. FVV streaming

A hatalmas adatmennyiség hálózaton való továbbítása nehézséget okoz, melynek kezeléséhez érdemes áttekinteni, hogy milyen videófolyamokat kell kezelni. Az átvitelt két nagy szakaszra bonthatjuk: kamera-képszintézis és képszintézis-felhasználó közötti szakasz. Az első szakaszban a kamerák és mélységérzékelők képeit kell továbbítani, a második szakaszban pedig a felhasználók által igényelt egyedi nézőpontokhoz tartozó képeket. Kulcsszerepet játszik tehát a képszintézis helye a FVV architektúrában (3. ábra), melynek alapján három modellt különböztethetünk meg [7]:

- (i) médiaszerver alapú,
- (ii) kliens alapú és
- (iii) elosztott (proxy) modellt.

A médiaszerver alapú képszintézis esetén a kamerák képeiből helyben történik a virtuális nézőpont előállítása, majd az egyedi videófolyamok a hagyományos módon, streaming protokollok (RTP/UDP, MPEG-TS) segítségével, kerül továbbításra. Annak érdekében, hogy a szerver tudja, mely egyedi nézőpontokat kell létrehozni, folyamatos vezérlőinformáció átvitelre van szükség a felhasználótól a szerver irányába. Mivel minden előállított videófolyam egyedi, multicast megoldás nem alkalmazható a forgalom csökkentésére. Nagyszámú felhasználó szintén gondot jelenthet, hiszen a képszintézis rendkívül erőforrásigényes feladat, ami túlterhelheti a szervert. A skálázhatósági problémák miatt a médiaszerver alapú képszintézis csupán alacsony felhasználószám esetén javasolt.

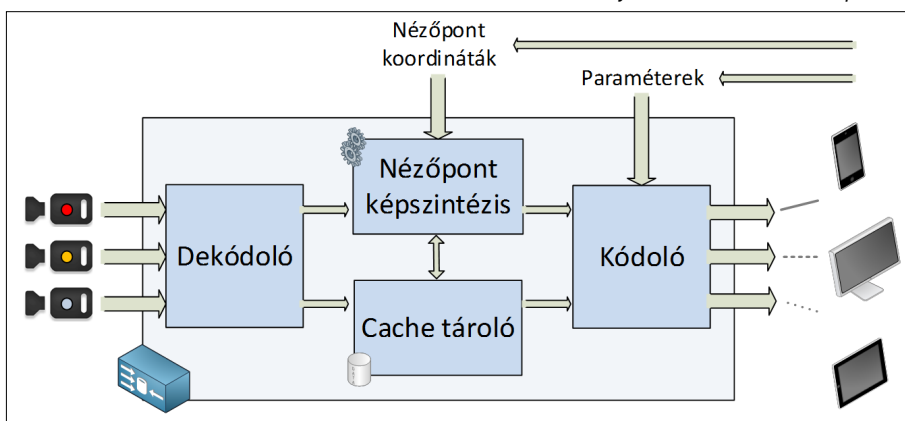
A második lehetőség a kliens alapú nézőpont előállítás, melynek feltétele, hogy a szükséges kameraképek a felhasználó eszközéhez eljussanak, ahol az igény szerinti, egyedi nézőpont előállításra kerül. A megoldás előnye, hogy nem kell vezérlőinformációt továbbítani a hálózaton, így a késleltetést is alacsonyabb szinten lehet tartani.

Hátrányt okozhat azonban a hatalmas adatforgalom, hiszen minden felhasználóhoz legalább két kameraképet kell eljuttatni, de adott esetben akár hármat vagy négyet is. Több kameraképre akkor lehet szükség, ha síkbeli kameraelrendezést alkalmazunk, ami egyben nagyobb mozgásteret ad a felhasználó nézőpontjának. A másik eset az új kameraképeket igénylő nézőpontváltás felgyorsítása esetén lehetséges. Ha meg tudjuk becsülni, mely kamerákra lesz szüksége a nézőnek a közeljövőben, akkor ezek előtöltésével hatékonyabban fedhető el a kameraváltások okozta megszakadások. A kliens alapú nézőpont előállítás egyik előnye, hogy a hálózaton a valós kameraképeket kell továbbítani, melyek esetén a multicast routing hatékony forgalomcsökkentést eredményezhet. Ez a módszer nagyszámú felhasználó esetén jó választás.

Végül, a harmadik lehetőség, ha elosztott módon szolgáljuk ki a felhasználókat. A nézőpont előállítás ebben az esetben több proxy szerveren történik. A valós kamerák képeit a proxy szerverekhez kell eljuttatni, majd onnan már az egyedi videófolyam továbbítása történik. Ez a módszer akkor hatékony, ha felhasználó eszköze nem rendelkezik elegendő számítási kapacitással a képszintézis elvégzésére, ugyanakkor a nagyszámú felhasználó kiszolgálása pedig skálázhatósági problémát jelentene a központ médiaszerver számára. A proxy szerver, illetve médiaszerver általános modelljét a 4. ábra mutatja.

A proxy szerver és kliens szerver alapú megoldásnál a multicast továbbítási mód mellett további forgalomcsökkentést tudunk elérni, ha főleg kameraképeket nem továbbítunk a képszintézis végző eszköznek, csupán azokat, melyekre valóban szüksége van, vagy szükséges lehet a közeljövőben.

4. ábra Proxy/média szerver felépítése





5. ábra Kamerakép és mélységkép

## 5. Nézőpont előállítás, képszintézis

Számos 3D technológia elérhető manapság, melyek közül a Light Field Rendering (LFR) [8] és Depth Image Based Rendering (DIBR) [9] technológiát érdemes kiemelni. A LFR technológia tisztán kép alapú képszintézis technológia, mely a kameraképek különböző módszerek alapján történő interpolálásával hozza létre a kívánt nézőponthoz tartozó képet.

DIBR esetén a színes kameraképek és mélységképek felhasználásával történik a virtuális nézőpont előállítása (5. ábra). 3D térre vetített pontokat a virtuális kamera képére vetítjük, így előállítva a virtuális nézőpontot. Ezt a 2D-3D, majd 3D-2D transzformációt „3D image warping” technikának nevezik a számítógépes képkalkotásban. Ha a szintetizált nézőpont sokkal távolabb helyezkedik el, mint az eredeti kamerakép, úgy képhibák jelenhetnek meg abból fakadóan, hogy a rendszernek nincs információja a tér azon pontjairól. Ilyen hibák keletkezhetnek akkor is, ha a mélységkép nem pontos, ez pedig gyakran előforduló hiba. A hibák elfedéséhez különböző algoritmusok léteznek [10].

## 6. Megjelenítés

A 3D tartalom megjelenítésére különböző eszközök léteznek, melyek közül a legegyszerűbb a hagyományos 2D kijelzők alkalmazása, annyi módosítással, hogy a megjelenített nézőpont a néző igényei szerint kerül megjelenítésre. Ezzel a módszerrel találkozhatunk például a számítógépes játékok, 3D modellező szoftverek esetén. Az LFR, illetve DIBR képszintetizáló algoritmusok segítségével azonban már élő, képalapú nézőpont előállítás valósítható meg. A 2D kijelző alkalmazásával a perspektivakövetés ugyan megvalósítható, de igazi mélységérzet mégsem adható vissza.

Valós mélységérzetet és szabad nézőpont választást támogató kijelzők közé tartoznak a virtuális sisakok. Annak ellenére, hogy ma már alacsony késleltetésű és nagy felbontású eszközök állnak rendelkezésre, viselésük kényelmetlen. Speciális igények esetén és 3D játékokra megfelelnek ezek a megjelenítők, de a hétköznapi médiafogyasztásra alkalmatlanok.

Szerencsére kezdenek megjelenni olyan kijelzők is, melyek nem igényelnek speciális szemüveget vagy si-

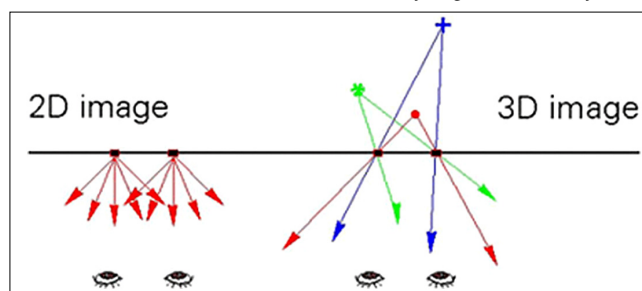
sakot, ugyanakkor a perspektivakövetés mellett mélységérzet is adnak. A Light Field kijelzők működésének alapját a fénysugarak irányszűrése adja. A hagyományos kijelző képpontjait elhagyó fénysugár a tér minden irányába haladhat, míg a 3D kijelzők esetén csupán megadott térszögben történhet a sugárzás, mint ahogy a magyar vonatkozású HoloVizio [11] megjelenítő is teszi. A módszer segítségével elérhető, hogy a megjelenített objektum olyan hatást adjon, mintha az a képernyő síkja előtt vagy mögött helyezkedne el, ahogy a 6. ábra szemléleti.

Az áttörést hozó kijelzők ma már megvásárolhatók, azonban még jelentős fejlődésen kell átesniük ahhoz, hogy a mindennapjaink részévé váljanak. Emellett azal is tisztában kell lennünk, hogy a kijelző önmagában semmit nem ér, ha nincs hozzá tartalom.

## 7. Összefoglalás

A 3D multimédia folyamatos és gyors fejlődésen megy keresztül napjainkban, melynek legújabb irányzata a Free Viewpoint Video szolgáltatás. Az új technológia lehetőséget ad a felhasználónak, hogy valós időben tetszőleges nézőpont alapján szemlélje a videótartalmat. A rendszer alapja, hogy több kamera által rögzítik a jelenetet, majd a kameraképekből az alkalmas képszintetizáló algoritmus segítségével tetszőleges nézőpont állítható elő. A szolgáltatás megvalósításához számos részfeladatot kell megoldani, a képrögzítés, tömörítés, átvitel képszintézis és megjelenítés területén. Ha a folyamatlánc minden eleme hibátlanul és kimagasló módon képes működni, csupán akkor várhatjuk, hogy a tartalomelőállítók is támogassák a technológiát, melytől valóban a nagy áttörést várunk a 3D videózás területén.

6. ábra Mélységérzet 3D kijelzőn



## Irodalom

- [1] Aljoscha Smolic,  
„3D video and free viewpoint video – From capture to display”, Pattern Recognition,  
Vol. 44, Issue 9, pp.1958–1968., September 2011.
- [2] T. Kobayashi, T. Fujii, T. Kimoto, M. Tanimoto,  
„Interpolation of Ray-Space Data by Adaptive Filtering”  
IS&T/SPIE Electronic Imaging, 2000.
- [3] P. Merkle, Y. Morvan, A. Smolic, D. Farin, K. Muller,  
T. Wiegand,  
„The effects of multiview depth video compression  
on multiview rendering”, Signal Processing:  
Image Communication, pp.73–88., 2009.
- [4] P. Merkle, A. Smolic, K. Mueller, T. Wiegand,  
„Efficient prediction structures for multiview video  
coding”, IEEE Transactions on Circuits and Systems  
for Video Technology, Special Issue on Multiview  
Video Coding and 3DTV, 2007.
- [5] ISO/IEC 23002-3:2007 Information technology –  
MPEG video technologies, Part 3: Representation of  
auxiliary video and supplemental information, 2007.
- [6] Guan-Ming Su, Yu-Chi Lai, Andres Kwasinski,  
Haohong Wang,  
„3D video communications: Challenges and  
opportunities”, Int. Journal of Com. Systems,  
Vol. 24/10, pp.1261–1281, October 2011.
- [7] Árpád Huszák,  
„Advanced Free Viewpoint Video  
Streaming Techniques”,  
Int. Journal on Multimedia Tools and Applications,  
Springer, DOI: 10.1007/s11042-015-3048-9, 2015.
- [8] M. Levoy, P. Hanrahan,  
„Light field rendering”,  
Computer Graphics, Proceedings. SIGGRAPH96,  
August 1996.
- [9] Christoph Fehn,  
„Depth-image-based rendering (DIBR), compression,  
and transmission for a new approach on 3D-TV”,  
Proc. of SPIE, Vol. 5291, Stereoscopic Displays and  
Virtual Reality Systems, pp.93–104., May 2004.
- [10] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester,  
Image repainting,  
Proc. of the ACM SIGGRAPH, 2000.
- [11] Holografika, HoloVizio,  
<http://www.holografika.com/>

## A szerzőről



**HUSZÁK ÁRPÁD** a Budapesti Műszaki és Gazdaságtudományi Egyetemen szerezte villamosmérnöki oklevelét 2003-ban, majd 2010-ben védte meg PhD fokozatát a BME Híradástechnikai Tanszékén, IP-alapú multimédia szolgáltatások témakörben. A Hálózati Rendszerek és Szolgáltatások Tanszékén (korábbi nevén Híradástechnikai Tanszékén) 2010-től adjunktusként dolgozik a Multimédia Hálózatok és Szolgáltatások Laboratórium (MEDIANETS) tagjaként. Kutatási területe a multimédia kommunikáció.

Számos munkája jelent meg az újfajta transzport protokollok (UDP-lite, DCCP, SCTP) média-átviteli teljesítményének vizsgálata témakörében. Munkája során a vezeték nélküli környezet, videó minőségre gyakorolt hatását kutatja. Az elmúlt néhány évben a 3D és Free Viewpoint Video átvitel okozta nehézségek kerültek kutatásának fókuszába.