

# Szakmai beszámoló

A Magyar Állami Eötvös Ösztöndíj pályázat keretében egy hónapot tölthettem az Amerikai Egyesült Államok egyik legrangosabb egyetemén, a Cornell University-n. Az egy hónap alatt lehetőségem volt bekapcsolódni a Cornell Computing and Information Science karon található számítógépes grafikával foglalkozó kutatócsoport munkájába. Igazán szerencsésnek érzem magam, hiszen a munkámat Kavita Bala, a szakma elismert képviselője, egyben a kar dékánja és a csoport vezetője felügyelte. Az ösztöndíjas időszak sajnos az eredeti terveimhez képest viszonylag rövid volt, mégis úgy érzem, hogy sikerült a lehető leghasznosabban eltölteni ezt a bő egy hónapot.

Amerikába érkezésemet követően azonnal fel tudtam venni a kapcsolatot az egyetem tanulmányi osztályával, így a kezdeti adminisztráció nagyon gördülékenyen zajlott. Mivel a kutatásom témáját sikerült előzetesen megvitatnunk, a munkatervemnek megfelelően augusztus első hetében meg is tudtam kezdeni az érdemi kutatómunkámat. A fogadó professzorom készségesen a rendelkezésemre állt, és segített beilleszkedni a csoportba, illetve összekötött a projekten dolgozó többi csoporttaggal. Teljes hozzáférést kaptam a kutatólabor erőforrásaihoz, és a kint tartózkodásom idejére biztosították számomra a munkámhoz szükséges összes eszközt.

## KUTATÁSI PROJEKT

A kutatásom céljával a gépi tanulási módszerek integrálását tűztük ki vizualizációs környezetekbe. Ezen belül is a neurális hálózaton alapuló 3D jelenetrekonstrúció és képszintézis (*neural rendering*) volt a témám. Augusztus első hetében az időm nagy részét a témával való ismerkedés és a projekthez kapcsolódó szakirodalom tanulmányozása tette ki. Az egyetemnek aktív előfizetése van a legtöbb rangos folyóiratra, így a kutatásomhoz kapcsolódó publikációkhoz gond nélkül hozzáfértem. Az elmúlt egy évben számtalan tudományos értekezés jelent meg a neurális hálózattal történő 3D jelenetrekonstrúció és képszintézis területén. Ezek közül azonban kevés foglalkozik azzal a kérdéssel, hogy hogyan lehet az áttetsző objektumokat (pl. üveg, műanyag, ...) megfelelően modellezni. A neurális hálózaton alapuló képszintézis egyik gyenge pontja, hogy az áttetsző objektumokat korlátozottan, csak meglehetősen rossz minőségben képes rekonstruálni. Ennek az az oka, hogy az áttetsző objektumok felületén általában különféle tükröződő és csillogó fényjelenségeket tapasztalhatunk, amelyek ráadásul a fényforrástól és a nézeti iránytól függően más és más hatást keltenek. A feladatot tovább nehezíti, hogy az átlátszó objektumoknak nincsen saját textúrájuk.

Helyette mindig a mögöttük látható háttér torzított képét láthatjuk, ami természetesen szintén a nézeti iránytól függ. A jelenleg elfogadott leghatékonyabb módszerek olyan megoldásokon alapulnak, amelyekhez speciális kamerarendszerre van szükség. Ezenfelül a javasolt módszerek többsége olyan további paraméterekre támaszkodik, amelyek méréséhez laboratóriumi körülményekre van szükség.

Jelen projekt keretében célunk egy olyan gépi tanuláson alapuló módszer kidolgozása, amely lehetővé teszi az átlátszó objektumok jó minőségű rekonstrukcióját a fenti erős megkötések nélkül. Mindezt úgy szeretnénk elérni, hogy a rekonstrukcióhoz csak olyan eszközökre legyen szükség, amelyek bárki számára könnyen hozzáférhetőek (pl. okostelefon), és amelyekkel a mérések akár otthoni körülmények között is egyszerűen elvégezhetőek (pl. 10-20 fotó készítése különböző irányokból a rekonstruálni kívánt objektumokról).

## **A JAVASOLT MÓDSZER KIDOLGOZÁSA**

Augusztus második hetében megkezdtük az általunk javasolt módszer kidolgozását. Ehhez első lépésként közelebbről is meg akartunk ismerkedni a jelenleg elfogadott technikákkal. Kiindulásként három közelmúltban megjelent publikációt vettünk elő [1,2,3].

Az [1] és [2] módszer átlátszó objektumokról alkotott képek alapján próbál mélységinformációt rekonstruálni, amelyet aztán autonóm rendszerekben alkalmaznak. Ilyen rendszer lehet például egy robotkar, amelynek a feladata az objektumok megragadása és a pozíciójuk megváltoztatása. Átlátszó objektumok esetén a mélységszenzorokból érkező elsődleges információ túl zajos a közvetlen felhasználáshoz. Az [1] és [2] módszerrel ezen próbálnak segíteni neurális hálózat alapú technikákkal. A feladat itt ugyan nem teljesen azonos az általunk megoldani kívánt problémával, ugyanakkor a szerzők megosztották a modell tanításához használt adathalmazt, amit mi is fel tudtunk használni a projekt kezdeti szakaszában.

A [3] cikk már közelebb áll a mi feladatunkhoz. A szerzők itt egy bonyolult modellt dolgoztak ki, amely több komponens együttműködését feltételezi. Az általuk javasolt módszer alkalmazhatóságát nehezíti az a tény, hogy a rekonstrukcióhoz a különböző nézeti irányokból készített fotók mellett egy környezetterképre<sup>1</sup>, valamint az üvegobjektum fizikai paramétereire is szükség van. Az eredményeik meggyőzőek. Az átlátszó objektumokról szintetizált képeik valóban nehezen megkülönböztethetőek a valódi képektől. A módszer hátránya a modell bonyolultsága, illetve a nehézkesen kivitelezhető mérések, amelyek szükségesek a rekonstrukcióhoz.

---

<sup>1</sup> environment map

A fenti módszerekkel ellentétben mi olyan megoldást szeretnénk nyújtani, amely egyszerűen, akár otthoni körülmények között is kivitelezhető méréseket igényel. Mögöttes modellnek a [4] cikk által javasolt reprezentációt választottuk. Itt a 3D jelenetet egy sűrűn összekötött neurális hálózat (MLP<sup>2</sup>) reprezentálja, amely képes a térbeli pontokhoz egy sűrűségértéket és egy nézeti irányfüggő színértéket rendelni. Ezekből az adatokból térfogat-vizualizációs technikákkal lehetővé válik bármilyen –, akár a hálózat tanításakor ismeretlen – nézeti irányból a képszintézis.

A kezdeti irodalomkutatást követően elkezdtük az általunk választott neurális hálózat alapú modell betanítását. Ehhez szükségünk volt egy olyan adathalmazra, amely megfelelő mennyiségű képet tartalmaz átlátszó objektumokról. A csoport a korábbi projektjeihez a Mitsuba<sup>3</sup> nevű renderelőt alkalmazta, így kézenfekvő volt, hogy ehhez a projekthez is ezt az eszközt hívjuk segítségül. Kezdetben egyszerűbb jelenetekkel kísérleteztünk, ahol összesen egy objektumot helyeztünk el egy virtuális asztallapon. A kiindulási modelltől azt vártuk, hogy nem lesz képes az átlátszó objektumokat megfelelően rekonstruálni. A kísérleteink ezt a feltételezést részben alátámasztották, azonban meg kell jegyezni, hogy nagyon egyszerű jelenetek esetén már a kezdeti modellel is képesek voltunk kielégítő eredményeket elérni.



1. ábra - Borospohár rekonstrukciója a kezdeti modellel. Balra látható az elvárt kimenet, középen a hálózat által szintetizált kép, jobbra pedig a diszparitás térkép.

Következő lépésként szeretnénk volna a modellbe beépíteni az RGB információn kívül a mélység információt is. Az ötletünk az volt, hogy *semi-supervised learning* segítségével tanítjuk a hálózatot. Ennek az a lényege, hogy az adatok egy részét plusz információval látjuk el, és így csatoljuk a hálózat bemenetére. Esetünkben ez a plusz információ az RGB képek mellé csatolt mélységkép volt. Először egy szintetikus adathalmazzal kísérleteztünk, amelyet a Mitsuba

<sup>2</sup> Multi-Layer Perceptron

<sup>3</sup> <https://www.mitsuba-renderer.org/>

renderelővel generáltunk. A kezdeti eredményeket látva úgy döntöttünk, hogy a kísérleteket megismételjük valós adatokon is. A valós adathalmaz elkészítéséhez egy Intel RealSense D415 kamerát használtunk, amellyel mind az RGB-, mind a mélységképeket rögzíteni tudtuk az általunk előkészített jelenetről. A jelenetben vegyesen szerepeltek átlátszó és átlátszatlan objektumok. Az Intel kamerából származó mélységinformáció átlátszó objektumok esetén meglehetősen zajos, ezért a valós adathalmazzal a modell rosszabb eredményeket produkált. Jelenleg azon dolgozunk, hogy a neurális hálózat felépítésének a változtatásával egy olyan robusztus modellt kapjunk, amely a mélységképeken egy automatikus zajszűrést hajt végre, és képes az átlátszó objektumok megjelenítését jelentősen javítani.

## ÖSSZEFOGLALÁS

A Cornell Egyetemen töltött egy hónap alatt lehetőségem volt bekapcsolódni a számítógépes grafika csoport munkájába egy számomra igazán érdekes és aktuális téma kutatásának a keretében. A választott témám a neurális hálózaton alapuló 3D jelenetrekonstrúció és képszintézis (*neural rendering*) volt. Ezen belül azzal foglalkoztam, hogy hogyan lehet a neurális hálózaton alapuló képszintézis minőségén javítani az átlátszó objektumok esetében. A projekten jelenleg is dolgozom a Cornell Egyetemmel együttműködésben. A kint megkezdett kutatás eredményeit szeretném a Ph.D. disszertációmba beépíteni, valamint a kinti kutatócsoporttal közösen publikálni.

**HIVATKOZÁSOK**

- [1] S. S. Sajjan *et al.*, “ClearGrasp: 3D Shape Estimation of Transparent Objects for Manipulation,” *arXiv:1910.02550 [cs, eess]*, Oct. 2019, Accessed: Aug. 21, 2021. [Online]. Available: <http://arxiv.org/abs/1910.02550>
- [2] M. Mousavi and R. Estrada, “SuperCaustics: Real-time, open-source simulation of transparent objects for deep learning applications,” *arXiv:2107.11008 [cs]*, Jul. 2021, Accessed: Aug. 23, 2021. [Online]. Available: <http://arxiv.org/abs/2107.11008>
- [3] Z. Li, Y.-Y. Yeh, and M. Chandraker, “Through the Looking Glass: Neural 3D Reconstruction of Transparent Shapes,” *arXiv:2004.10904 [cs]*, Jul. 2020, Accessed: Sep. 13, 2021. [Online]. Available: <http://arxiv.org/abs/2004.10904>
- [4] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” *arXiv:2003.08934 [cs]*, Aug. 2020, Accessed: Aug. 18, 2021. [Online]. Available: <http://arxiv.org/abs/2003.08934>