Posudek oponenta diplomové práce

Autor práce: *Bc. David Otgonsuren Rico*

Název práce: *3D Human Model Reconstruction and Automatic Rigging From a Monocular Video*

Oponent práce: *prof. Ing. Daniel Sýkora, Ph.D.*, ČVUT FEL (K13139)

Cílem diplomové práce *Davida Otgonsuren Rico* bylo vytvořit věrný otexturovaný 3D model oblečené lidské postavy (staticky stojící), okolo které krouží kamera a pořizuje videozáznam. Výsledný 3D model mělo být dále možné rozpohybovat s využitím existujících technik skeletální animace. Student měl prostudovat nejlepší známé metody pro obecnou 3D rekonstrukci z videa, některé z nich aplikovat na vstupní sekvence a porovnat kvalitu výsledných 3D modelů. Dále se měl pokusit tyto modely rozpohybovat s využitím existujících technik pro skeletální animaci. Na závěr měl být implementován kompletní řetězec operací, který by ze vstupního videa vytvořil animovatelný 3D model.

Student pro srovnání vybral metody [3, 5, 6], které řeší obecný problém 3D rekonstrukce z videosekvencí a existují pro ně referenční implementace. Mrzelo mě, že nebyla porovnána či alespoň diskutována specializovaná metoda [4], na kterou jsem studenta s předstihem upozorňoval a která měla dle mého názoru výborné předpoklady v zadaném úkolu uspět.

Na základě výsledků srovnání zvolil student pro rekonstrukční část řetězce metodu NeuS-Facto použitou v nástroji SDFStudio [6]. Pro její správné fungování je nezbytné poskytnout i segmentační masky postavy. Ty student (po provedeném srovnání) vygeneroval přístupem popsaným v článku [1]. Vlastní animace 3D modelů byla realizována s využitím metody [2]. Podařilo se tak implementovat kompletní řetězec. Jeho funkčnost byla následně ověřena na ukázkové sadě pěti testovacích videosekvencí. Lze tedy konstatovat, že zadání práce bylo splněno. Kvalita předložených výsledků ale bohužel nepůsobí úplně přesvědčivým dojmem ve srovnání s přístupem popsaným v článku [4]. Při pohledu do přiložených zdrojových kódů jsem byl navíc zklamán faktem, že se v zásadě jedná jen o sekvenci volání existujících implementací. Z textu práce nebylo zřejmé, které části student skutečně sám implementoval a které jen převzal. Vzhledem k tomu, že se jedná o práci magisterskou v oboru umělá inteligence, předpokládal bych, že alespoň část jinak komplikovaného řetězce bude student implementovat samostatně.

Ani vlastní text práce mě nějak výrazně nepotěšil. Působí spíše jako rozpracovaný koncept. Na mnoha místech je veden formou výčtu bodů a obsahuje četné překlepy. Od počátku není úplně zřejmé, jaký specifický problém se vlastně řeší. Úvodní část je velmi stručná a jen krátce čtenáře informuje o metodách, které byly v řetězci použity. Postrádal jsem detailnější rozbor současného stavu poznání v dané problematice (viz zmíněná chybějící reference na článek [4]). Část věnující se animaci mi přišla velmi stručná. Chybí ukázky dosažených výsledků, které nejsou ani součástí oficiální přílohy. Podařilo se mi je dohledat až na externím odkazu na Google Drive.

Přestože je téma prace *Davida Otgonsuren Rico* zajímavé a komplexní, nejsem úplně spokojen s mírou zapojení v implementační části a také s podobu výsledného textu. Práci však doporučuji k obhajobě a s přihlédnutím k výše uvedeným nedostatkům ji hodnotím stupněm **D-uspokojivě**.

V Praze, 13. června 2024 *prof. Ing. Daniel Sýkora, Ph.D.*

Literatura

[1] Cheng, H. K.; Schwing, A. G.: XMem: Long-Term Video Object Segmentation with an Atkinson-Shiffrin Memory Model. In *Proceedings of European Conference on Computer Vision*, 2022, s. 640–658.

[2] Li, P.; Aberman, K.; Hanocka, R.; aj.: Learning Skeletal Articulations with Neural Blend Shapes. *ACM Transactions on Graphics*, ročník 40, č. 4, 2021: čl. 130.

[3] Li, Z.; Müller, T.; Evans, A.; aj.: Neuralangelo: High-Fidelity Neural Surface Reconstruction. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2023, s. 8456–8465.

[4] Tang, S.; Wang, G.; Ran, Q.; aj.: High-Resolution Volumetric Reconstruction for Clothed Humans. *ACM Transactions on Graphics*, ročník 42, č. 5, 2023: čl. 170.

[5] Wang, P.; Liu, L.; Liu, Y.; aj.: NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *Advances in Neural Information Processing Systems*, 2021, s. 27171–27183.

[6] Yu, Z.; Chen, A.; Antic, B.; aj.: SDFStudio: A Unified Framework for Surface Reconstruction. 2022.