디지털휴먼엔터테인먼트특론 과제 1

120220210 고재현

2023년 3월 8일

1 개요

1.1 목적

- 최근 대두되는 디지털 휴먼 관련 기반 기술들을 조사한다.
- 해당 기반 기술을 구현하기 위해 핵심적인 요소들에 대해 고려해 본다.

1.2 요소 기술 및 최근 합성 기술의 동향

한유아와 같은 디지털 휴먼을 만들기 위해서는 얼굴 합성 기술, 모션 합성 기술, 음성 합성 기술이 필요하다. 사실적인 표현을 위한 캐릭터 모델링 및 렌더링 기술도 중요하지만, 인공지능 기반의 기술들만 서술하도록 한다. 최근 대두되는 생성 모델은 크게 두 가지의 흐름으로 분류할 수 있다. normalizing flow와 LM과 Transformer를 결합한 모델이 그것이다. normalizing flow는 그 이름에서 알 수 있듯이, 데이터의 분포를 정규화하는 방식으로 데이터를 생성한다. 학습 속도는 느리지만, inference 속도는 적은 수의 point를 생성하는 데에는 충분하다[1]. LM 기반의 생성 모델은 고품질의 Tokenizer 및 Transformer 기반 예측 모델의 발전으로 최근 Vall-E[2]와 같이 빠른 속도의 학습 및 생성이 가능하다.

1.3 요소 기술별 동향

1.3.1 얼굴 합성 기술

얼굴 합성 기술은 주로 키포인트(랜드마크 등)를 생성한 뒤, 해당 키포인트를 기반으로 모델링을 덮어씌우는 방식으로 수행[3]되거나, 원본 영상을 변형하여 각 영상 프레임을 생성하는 방식[4]으로 수행된다. 스마일게이트 AI Media Team의 블로그 자료를 보면, 이와 관련하여 원본 영상을 변형하여 각 영상 프레임을 생성하는 방식의 실시간 처리를 시도한 바가 있다. 해당 방식은

2 디지털 휴먼 모델의 발전 방향

참고 문헌

[1] H.-K. Song, S. H. Woo, J. Lee, S. Yang, H. Cho, Y. Lee, D. Choi, and K.-w. Kim, "Talking face generation with multilingual tts," 2022.

[2] Z. Zhang, L. Zhou, C. Wang, S. Chen, Y. Wu, S. Liu, Z. Chen, Y. Liu, H. Wang, J. Li, L. He, S. Zhao, and F. Wei, "Speak foreign languages with your own voice: Cross-lingual neural codec language modeling," 2023.

[3] A. Richard, M. Zollhöfer, Y. Wen, F. D. la Torre, and Y. Sheikh, "Meshtalk: 3d face animation from speech using cross-modality disentanglement," *CoRR*, vol. abs/2104.08223, 2021.

[4] Y. Zhou, D. Li, X. Han, E. Kalogerakis, E. Shechtman, and J. Echevarria, "Makeittalk: Speaker-aware talking head animation," *CoRR*, vol. abs/2004.12992, 2020.