Simpósio de Integração Acadêmica
"Das Montanhas de Minas ao Oceano: Os Caminhos da Ciência para um Futuro Sustentável"
SIA UFV 2025

Universidade Federal de Viçosa

Geração de Ambientes Sintéticos para avatares de Linguagem de Sinais

[1]Gabriel Moreira Marques, [2]Thiago Luange Gomes, [3]Michel Melo da Silva

ODS 10 - Redução das Desigualdades

Pesquisa

Introdução

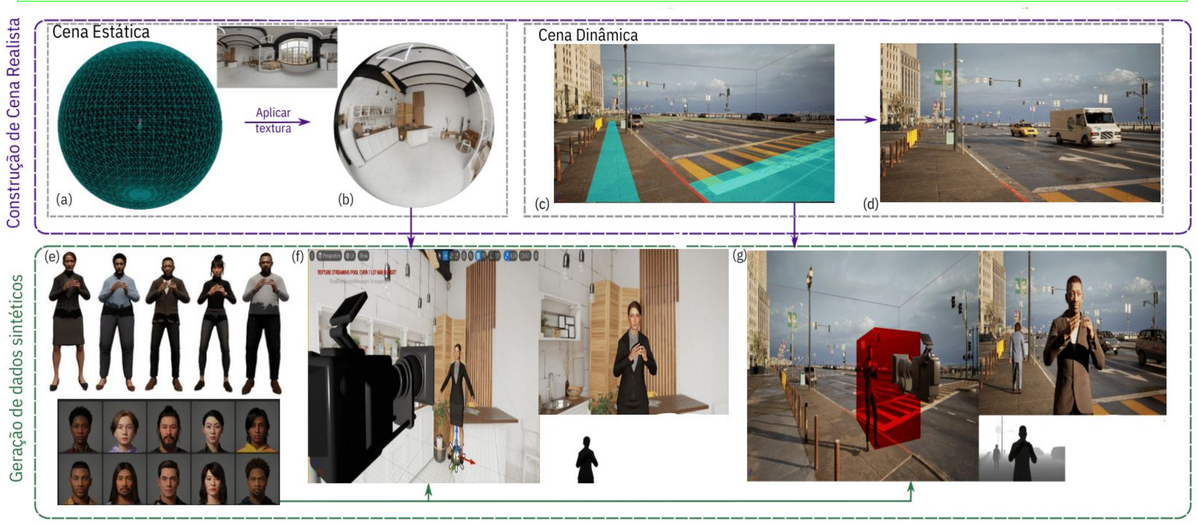
A crescente demanda por dados visuais e as limitações logísticas e éticas da coleta real tornam os dados sintéticos uma alternativa viável. Eles permitem controle sobre iluminação, enquadramento, roupas e cenários, reduzindo custos e riscos de privacidade. Para tarefas sensíveis ao visual, como reconhecimento de gestos em línguas de sinais, é essencial que avatares e cenários alcancem realismo e coerência contextual para preservar sinais finos de movimento.

Neste trabalho propomos um pipeline automatizado, baseado em Unreal Engine 5 e MetaHumans, para replicar movimentos humanos a partir de vídeos de entrada e renderizá-los em avatares de alta fidelidade, gerando vídeos RGB, mapas de profundidade, posições das juntas e parâmetros de câmera quadro a quadro.

Objetivos

Desenvolver e validar um pipeline automatizado para gerar dados sintéticos fotorealistas de avatares que reproduzem movimentos extraídos de vídeos, visando aplicações em reconhecimento de gestos e línguas de sinais.

Material e Métodos ou Metodologia



O pipeline opera em duas etapas integradas: (i) geração do cenário — estático (SkySphere texturizado com imagens HDRi ou dinâmico (City Sample com tráfego e pedestres) — e (ii) criação e animação de um avatar fotorrealista (MetaHuman). O avatar é parametrizado automaticamente (aparência e vestuário), posicionado na cena e animado a partir da estimativa de pose extraída do vídeo de entrada. A câmera é posicionada e verificada quanto a oclusões (reposicionamento automático quando necessário). A renderização quadro-a-quadro produz sequências RGB, mapas de profundidade e metadados de câmera sincronizados.

Apoio Financeiro

CAPES
FAPEMIG 40 ANOS
CNPq

Resultados e/ou Ações Desenvolvidas

Dados sintetizados utilizando o fluxo de trabalho proposto a partir de um vídeo de entrada (primeira linha), retratando diversidade em cena, avatar, roupas e fundo (da segunda à quinta linha), utilizando abordagens de cena estática (esquerda) e dinâmica (direita). A primeira coluna de cada abordagem de cena representa a profundidade do quadro mais à esquerda.

Conclusões

O pipeline converte vídeos em conjuntos de dados sintéticos anotados e visualmente plausíveis, combinando MetaHumans com cenas estáticas (SkySphere+HDRi) e dinâmicas (City Sample). Os MetaHumans oferecem alta fidelidade e esqueleto completo e, junto ao City Sample Crowds, possibilitam ampla variedade de roupas e aparências. A partir de uma única gravação gera-se múltiplas amostras sincronizadas (RGB, depth, joints, parâmetros de câmera), ampliando a base para detecção e interpretação de sinais em línguas de sinais. Apesar de desafios em harmonização de iluminação e do custo computacional das cenas dinâmicas, o método equilibra bem realismo e escalabilidade.

Bibliografia

EPIC GAMES. Unreal Engine. Versão 5.2. [S.l.], 2024. Disponível em: https://www.unrealengine.com/. Acesso em: 01 out. 2025.

AKADA, H.; et al. Unrealego: A new dataset for robust egocentric 3D human motion capture. In: EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 2022. p. 1–17.

LI, Y.; et al. MatrixCity: A large-scale city dataset for city-scale neural rendering and beyond. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV), 2023. p. 3205–3215.

BLACK, M. J.; et al. BEDLAM: A synthetic dataset of bodies exhibiting detailed lifelike animated motion. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2023. p. 8726–8737.