

Um estudo sobre reconstrução e transferência de movimentos de Pessoas reais para Humanos virtuais

Eduardo S. Cardoso¹, Soraia R. Musse¹

¹Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

Abstract. *This work studies the transfer of facial expressions from real individuals to digital avatars, using advanced technologies such as Unreal Engine 5 and the MetaHuman platform. The research focuses on accurately capturing facial expressions through photogrammetry and their subsequent modeling into digital avatars, exploring the effectiveness of Action Units (AUs) to represent basic emotions such as anger, happiness, fear, disgust, sadness and surprise. Detailed analyzes were carried out to compare the intensity of facial expressions between real and virtual models, evaluating the smoothing of expressions when transferred to the digital environment. The investigation revealed significant variations in the fidelity of facial expressions, with MetaHuman technology showing superior ability to maintain the intensity and detail of the original expressions.*

Resumo. *Este trabalho estuda a transferência de expressões faciais de indivíduos reais para avatares digitais, utilizando tecnologias avançadas como Unreal Engine 5 e a plataforma MetaHuman. A pesquisa concentra-se na captura precisa de expressões faciais através de fotogrametria e sua subsequente modelagem em avatares digitais, explorando a eficácia das Action Units (AUs) para representar emoções básicas como raiva, felicidade, medo, nojo, tristeza e surpresa. Foram realizadas análises detalhadas para comparar a intensidade das expressões faciais entre os modelos reais e os virtuais, avaliando a suavização das expressões ao serem transferidas para o ambiente digital. A investigação revelou variações significativas na fidelidade das expressões faciais, com a tecnologia MetaHuman mostrando capacidade superior em manter a intensidade e detalhes das expressões originais.*

1. Introdução

Este estudo investiga a transferência de expressões faciais de indivíduos reais para avatares digitais, empregando tecnologias avançadas como a *Unreal Engine 5* e a plataforma *MetaHuman*. O principal objetivo deste estudo é facilitar e desenvolver métodos para analisar algo tão subjetivo quanto às expressões humanas, utilizando um modelo que possa capturar e reproduzir com precisão estas expressões em ambientes virtuais. Para alcançar isso, o trabalho emprega técnicas de fotogrametria para captura detalhada das características faciais, seguida pela aplicação das *Action Units* (AUs) de [Paul Ekman 1978] para categorizar e quantificar as expressões emocionais básicas, incluindo felicidade, raiva, medo, nojo, tristeza e surpresa.

Através da utilização do *MetaHuman*, uma plataforma conhecida por gerar modelos 3D fotorrealistas, este estudo analisou a precisão com que as expressões faciais são preservadas ao serem transferidas do real para o digital. Foram realizadas comparações

entre as intensidades das AUs em faces reais e suas contrapartes virtuais, identificando o grau de suavização ou perda de intensidade resultante do processo de digitalização. Este aspecto é crucial, pois a fidelidade com que as emoções são representadas em avatares digitais tem implicações diretas para áreas como a realidade virtual, jogos e simulações interativas, onde a autenticidade emocional é essencial.

A análise incluiu também uma avaliação comparativa com estudos anteriores para destacar as melhorias ou desafios remanescentes na tecnologia de captura e animação facial. O estudo mostrou que, apesar das inerentes dificuldades em capturar a sutileza das expressões humanas, as técnicas atuais permitem uma representação consideravelmente precisa, embora algumas expressões ainda experimentem uma certa suavização quando convertidas para o formato digital.

Em resumo, este trabalho não apenas avança a compreensão técnica e teórica da representação digital de expressões humanas, mas também desenvolve ainda mais a metodologia desenvolvida por [Vitor Miguel Xavier Peres and Musse. 2023] para avaliar e comparar expressões faciais reais e virtuais. Esta abordagem permite uma análise detalhada da fidelidade com que as emoções são reproduzidas em avatares digitais, crucial para aplicações em realidade virtual, jogos e simulações interativas. A metodologia desenvolvida busca garantir que as nuances das emoções humanas sejam capturadas e representadas eficazmente, melhorando significativamente a interação humano-computador em ambientes digitais.

2. Fundamentação e Conceitos

Esta seção é essencial para uma base sólida para compreender as tecnologias e metodologias empregadas neste trabalho. Detalhando como cada componente tecnológico contribui para a representação fidedigna de expressões faciais em avatares digitais.

2.1. Unreal Engine 5

A *Unreal Engine 5* é uma das principais plataformas de desenvolvimento de jogos e simulações interativas do mercado. Ela se destaca pela capacidade de criar experiências visuais impressionantes, oferecendo ferramentas avançadas de renderização e recursos inovadores, como a criação de humanos virtuais chamados de *MetaHuman*.

2.2. MetaHuman

Os *MetaHumans* são personagens digitais altamente realistas e expressivos criados pela Epic Games em colaboração com a empresa 3Lateral ¹. Esses modelos são construídos a partir de uma combinação de tecnologias avançadas, incluindo captura de expressões faciais, escaneamento 3D e processamento de dados por meio de algoritmos de inteligência artificial. O objetivo é alcançar um novo nível de fidelidade visual e de interação para personagens digitais em jogos, filmes, entretenimento e outras áreas.

2.3. Nível de detalhe ou LOD

O Nível de Detalhe (LOD - *Level of Detail*) na *Unreal Engine 5* é crucial para o desempenho e a qualidade visual dos *MetaHumans*. Ele permite que a engine gerencie diferentes

¹<https://www.3lateral.com/>

versões de um modelo, cada uma com níveis variados de complexidade geométrica e de textura. Isso é essencial para otimizar o desempenho, garantindo que, dependendo da distância do personagem ao observador, a *engine* carregue a versão mais adequada em termos de detalhes, preservando a qualidade visual sem sobrecarregar o hardware.

2.4. Fotogrametria

A fotogrametria é uma técnica utilizada para capturar objetos do mundo real e convertê-los em modelos digitais tridimensionais. Isso é feito por meio de fotografias tiradas de diferentes ângulos, que são processadas por software especializado para reconstruir a forma e as características do objeto. Na criação de personagens como os *MetaHumans*, a fotogrametria pode ser usada para capturar nuances faciais, expressões e detalhes anatômicos com precisão, permitindo uma fidelidade visual impressionante nos modelos digitais.

3. Trabalhos relacionados

No presente trabalho, será adotado o conceito de *Action Units* (AU) desenvolvido por [Paul Ekman 1978] como parte do Sistema de Codificação de Ação Facial (*Facial Action Coding System*, FACS). O FACS descreve 46 unidades de ação distintas que são fundamentais para capturar os movimentos da face humana. Muitos pesquisadores utilizam essas *Action Units* para uma análise concreta e analítica das expressões faciais humanas.

[Barrett et al. 2019] por exemplo, revisou a literatura explorando as expressões faciais comuns associadas a seis categorias emocionais: raiva, nojo, medo, felicidade, tristeza e surpresa. Seus estudos revelam como os rostos se movem durante diferentes estados emocionais, realizando avaliações sobre quais emoções são inferidas através dos movimentos faciais.

[Mäkäräinen et al. 2014] e colaboradores discutem a transferência de expressões de humanos reais para avatares virtuais, sugerindo que essa transição pode resultar numa suavização das expressões faciais. A Simplificação dessas expressões ao migrá-las do domínio real para o CG (computação gráfica) também proporciona aos artistas maior liberdade para aprimorar as expressões de acordo com o estudo.

[Vitor Miguel Xavier Peres and Musse. 2023] investigou a transferência da felicidade genuína de atores reais para avatares virtuais, examinando tanto expressões faciais posadas quanto espontâneas em rostos reais e virtuais.

Em [Mavadati et al. 2013], o artigo estabelece protocolos para medição automática da intensidade das *Action Units*, detalhando o processo de acesso ao banco de dados para pesquisas em visão computacional, aprendizado de máquina e ciências comportamentais e afetivas. E em conjunto com [Ghayoumi and Bansal 2016] desenvolveu uma pesquisa para focar em mapear as principais características geométricas envolvidas nas expressões faciais a um subconjunto específico de AUs, que são únicas para determinadas expressões faciais.

Neste estudo, utilizaremos o conceito de *Action Units* em conjunto com as seis emoções básicas e suas expressões correspondentes para analisar a fidelidade das expressões em humanos virtuais em comparação com humanos reais que servem como modelo para o estudo. Os resultados obtidos serão comparados com os estudos de [Mäkäräinen et al. 2014] sobre a suavização das expressões em ambientes virtuais, bem

como com as descobertas de [Vitor Miguel Xavier Peres and Musse. 2023] em relação às expressões de felicidade. A próxima seção apresenta mais detalhes do *MetaHuman* que foi a tecnologia usada neste trabalho.

4. MetaHuman

A tecnologia *MetaHuman*², desenvolvida pela Epic Games em 2021, no âmbito do motor gráfico *Unreal Engine 5*, tem como propósito a criação e desenvolvimento de modelos 3D humanos de alta fidelidade. Essa tecnologia abrange desde a construção de modelos 3D baseados em características reais ou fictícias até animações realistas, as quais têm aplicabilidade em diversos campos, como cinema, educação, inteligência artificial e uma ampla gama de outras áreas que se beneficiam da utilização de personagens virtuais.

O *MetaHuman* oferece um excelente controle de escalabilidade e performance, apresentando diferentes níveis de detalhamento (LODs)³. O nível de detalhe controla a complexidade da malha 3D do objeto e é determinado pela proximidade da câmera em relação ao objeto no ambiente virtual.

Essa variação de detalhamento é também influenciada pela plataforma operacional utilizada. Por exemplo, a Tabela 1 [lod] especifica plataformas disponíveis para o uso do *MetaHuman*, sendo o PC o de maior qualidade, suportando recursos como rastreamento de raios e texturas de até 8192 *pixels* de resolução. No entanto, as especificações de nível de detalhe para iOS e Android vão somente até o nível 3. Os níveis de LOD são detalhados na Tabela 2[lod], dividida de 0 a 7, sendo 0 o de maior qualidade gráfica e complexidade. As discrepâncias entre as qualidades podem ser visualizadas na Figura 1. O controle de nível de detalhe é sincronizado entre os diferentes elementos que compõem um *MetaHuman* (cabeça, cabelo, roupas etc.), garantindo uniformidade na qualidade, mesmo que possuam diferentes números de LODs e diferentes distâncias da câmera principal.

O cabelo no *MetaHuman* apresenta uma grande variedade de formas e níveis de detalhes, abrangendo desde penteados até fios individuais na pele humana para um maior realismo. Esses tipos de cabelos são armazenados como *Groom Assets*⁴ e podem ser configurados individualmente no editor específico de *Groom Assets*. Dado que os cabelos podem ser compostos por diferentes tipos de geometria, seja fios individuais ou malhas de baixa resolução, é crucial configurá-los para exibição nos LODs ideais. Essa configuração está exemplificada na Figura 2⁵.

Tabela 1. Nível de detalhe por plataforma

Plataforma	Ray Tracing Support	Hair	Best LOD	Max Texture Size
PC (Epic and Cinematic Quality)	Yes	Strands and Cards	0	8192
PC (Medium and Lower Quality)	No	Strands and Cards	0	8192
Mac	No	Cards	0	8192
iOS/Android	No	Cards	3	2048

²<https://www.unrealengine.com/en-US/metahuman>

³<https://dev.epicgames.com/documentation/es-mx/metahuman/technical-specifications-for-metahumans/metahuman-lods>

⁴<https://docs.unrealengine.com/4.27/en-US/WorkingWithContent/Hair/GroomAssetEditor/>

⁵<https://dev.epicgames.com/documentation/es-mx/metahuman/technical-specifications-for-metahumans/metahuman-lods>

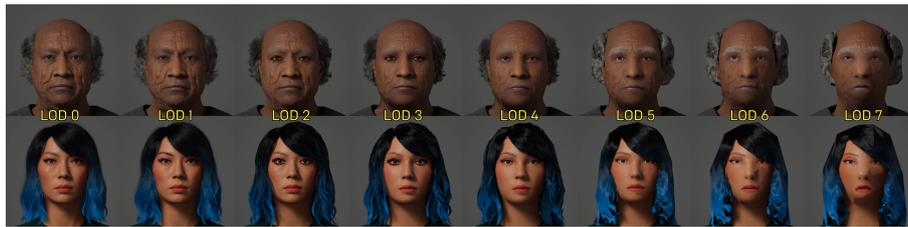


Figura 1. Diferentes níveis de detalhe entre os LODs 0 a 7.

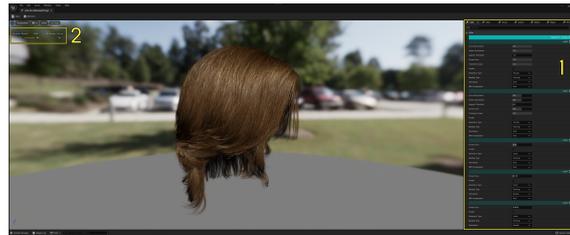


Figura 2. Aparência de um cabelo de *MetaHuman* no *Groom Asset Editor*. Painel LOD à direita (1). E caixa de informações à esquerda (2) fornecendo informações sobre o LOD atual e o número de fios relativos ao tamanho atual da tela.

4.1. Explorando alternativas ao *MetaHuman* para criação de modelos humanos 3D

A diversidade de ferramentas disponíveis para a criação de modelos tridimensionais humanos oferece uma gama de opções além do *MetaHuman* da *Unreal Engine 5*. Essas alternativas proporcionam flexibilidade e recursos exclusivos para artistas e desenvolvedores que buscam criar personagens 3D realistas e personalizados. Abaixo, são listadas algumas tecnologias que podem ser usadas como alternativas.

- O Blender se destaca como uma opção de software de código aberto que oferece um conjunto robusto de ferramentas para modelagem, escultura, texturização e animação. Sua comunidade ativa de usuários, juntamente com inúmeros tutoriais e recursos gratuitos, torna-o uma escolha popular para a criação de personagens humanos detalhados.
- Para aqueles que buscam um foco preciso na escultura digital, o ZBrush é uma referência. Reconhecido por sua interface intuitiva e poderosas ferramentas de escultura, este software é amplamente utilizado na indústria para esculpir detalhes intrincados em modelos 3D.
- O Reallusion CC3+ oferece uma abordagem simplificada para criar personagens 3D. Sua ênfase na facilidade de uso e na geração rápida de modelos humanos faz dele uma opção atraente, especialmente para iniciantes ou para projetos que demandam prazos mais apertados.
- Voltado para a criação de personagens humanos de forma acessível, o MakeHuman é um software de código aberto projetado especificamente para modelagem de personagens. Ele oferece uma variedade de opções personalizáveis para criar modelos humanos com diferentes características físicas.

Também no estudo [Vitor Miguel Xavier Peres and Musse. 2023], foram utilizadas três abordagens distintas para transformar expressões faciais em versões virtuais:

Tabela 2. Detalhamento por Nível de Qualidade

Level of Detail	0	1	2	3	4	5	6	7
Head								
Vertices	24000	12000	6000	2500	1300	560	270	130
Blendshapes	669	-	-	-	-	-	-	-
Joints	713	529	397	283	84	70	41	26
Skin Influences	12	12	12	8	8	8	4	4
Animated Maps	Yes	Yes	No	No	No	No	No	No
Body								
Body Vertices	30500	7600	3350	1507	-	-	-	-
Correctives	Yes	Yes	No	No	-	-	-	-
Skin Influences	8	8	8	4	-	-	-	-
Hair								
Hair Style Strands	50000	25000	-	-	-	-	-	-
Hair Style Card Vertices	30000	15000	10000	3000	1500	-	-	-
Hair Style Mesh Vertices	-	-	-	-	-	500	250	100
Facial Hair Strands	10000	5000	-	-	-	-	-	-
Facial Hair Card Vertices	15000	7000	3000	1000	500	-	-	-
Physics	Yes	Yes	Yes	No	No	No	No	No

- Deep3D [Deng et al. 2019], modelo baseado em aprendizado profundo desenvolvido para reconstrução tridimensional de objetos a partir de imagens 2D. Utiliza redes neurais profundas para aprender a mapear imagens bidimensionais em representações tridimensionais, explorando características complexas para reconstruir formas e estruturas.
- Emoca [Danecek et al. 2022] utiliza a *"perceptual emotion consistency loss"* (perda de consistência emocional perceptiva) como uma métrica inovadora no processo de reconstrução facial 3D para manter e validar a fidelidade emocional das expressões capturadas nas imagens de entrada. A *"perceptual emotion consistency loss"* é baseada no princípio de que as expressões faciais, quando reconstruídas em um modelo 3D a partir de uma fotografia 2D, devem preservar não apenas as características geométricas visíveis (como a forma dos olhos, boca e nariz), mas também a intensidade e nuances das emoções expressas.
- Deca [Feng et al. 2021] utiliza um mapa de deslocamento UV ⁶, gerado a partir de uma representação de baixa dimensão que combina parâmetros de detalhes específicos do sujeito e parâmetros genéricos de expressão. Esta abordagem permite a síntese de rugas realistas e específicas da pessoa ao controlar os parâmetros de expressão, mantendo os detalhes específicos da pessoa inalterados, assim gerando um modelo 3D detalhando.

Cada uma dessas ferramentas apresenta suas próprias vantagens e limitações. A escolha entre elas muitas vezes depende das necessidades do projeto, das habilidades do usuário e do estilo de criação desejado. Ao explorar essas alternativas, os artistas têm a

⁶É uma técnica usada em gráficos 3D para adicionar detalhes de alta resolução a um modelo 3D com uma malha de polígonos relativamente simples.

oportunidade de encontrar a ferramenta que melhor se adapta às suas preferências e fluxos de trabalho, permitindo a criação de modelos 3D de humanos expressivos e impactantes.

O *MetaHuman* se destaca atualmente por oferecer um ambiente mais avançado em comparação a essas outras ferramentas. Ele combina uma interface intuitiva com resultados de alta qualidade e complexidade. O *MetaHuman* permite abranger todas as etapas do processo, desde a modelagem até a animação de humanos 3D, integrando os modelos de humanos virtuais de forma eficaz com o poderoso motor gráfico da *Unreal Engine 5*. Por este motivo sendo o objeto de estudo do presente trabalho e de comparação com o estudo [Vitor Miguel Xavier Peres and Musse. 2023] na Seção 6.

5. Metodologia para Geração e Coleta de Dados

Para este estudo, devido à complexidade do processo e escopo do trabalho, foi realizado o processo de criação de humanos virtuais em apenas três indivíduos distintos, sendo um indivíduo masculino e dois femininos. Esses participantes passaram por todas as etapas de modelagem abordadas nesta seção.

As etapas seguidas para a geração de modelos no *MetaHuman* capazes de gerar expressões se baseiam em dois pontos principais: a geração do modelo 3D e a captura de expressões, com a subsequente transferência dessas expressões para o humano virtual.

Para o processo de análise da fidelidade das expressões, utilizamos o OpenFace, uma ferramenta de código aberto para a análise de movimentos faciais. O OpenFace permite a detecção de AUs e a medição da intensidade das expressões faciais, facilitando a avaliação da precisão e realismo das expressões geradas nos modelos virtuais. O processo do uso do OpenFace pode ser observado na Seção 5.2.

5.1. Modelos e expressões virtuais

O processo de criação de um modelo 3D do *MetaHuman* envolve várias etapas bem definidas, conforme ilustrado na Figura 4. A primeira etapa consiste na produção de um modelo 3D do rosto do indivíduo por meio da técnica de fotogrametria, conforme mostrado nas etapas 1 e 2 da Figura 4. Os aplicativos de fotogrametria disponíveis para celular em sua maioria requerem um iPhone devido à necessidade da câmera de profundidade. Neste trabalho, utilizou-se o PolyScan 3D ⁷, para a fotogrametria do rosto de uma pessoa, conforme mostrado na etapa 1.

Após a geração do modelo 3D inicial, o próximo passo é a limpeza e refinamento do modelo usando software especializado, como o Blender, para garantir precisão nos detalhes faciais. Esta etapa de limpeza é crucial para remover imperfeições e preparar o modelo para a integração na *Unreal Engine 5*.

Dentro da *Unreal Engine 5*, é gerado o modelo base do *MetaHuman*, que pode ser visto na etapa 3 da Figura 4. A partir desse ponto, ocorre a customização manual usando a interface da *engine* do *MetaHuman* para que ele se adeque ao indivíduo alvo, como pode ser observado na Figura 3. Essa customização inclui ajustes na cor da pele, características faciais e outros detalhes específicos para garantir um alto grau de semelhança e realismo, resultando no *MetaHuman* finalizado, como ilustrado na etapa 4 da Figura 4.

⁷<https://poly.cam/>

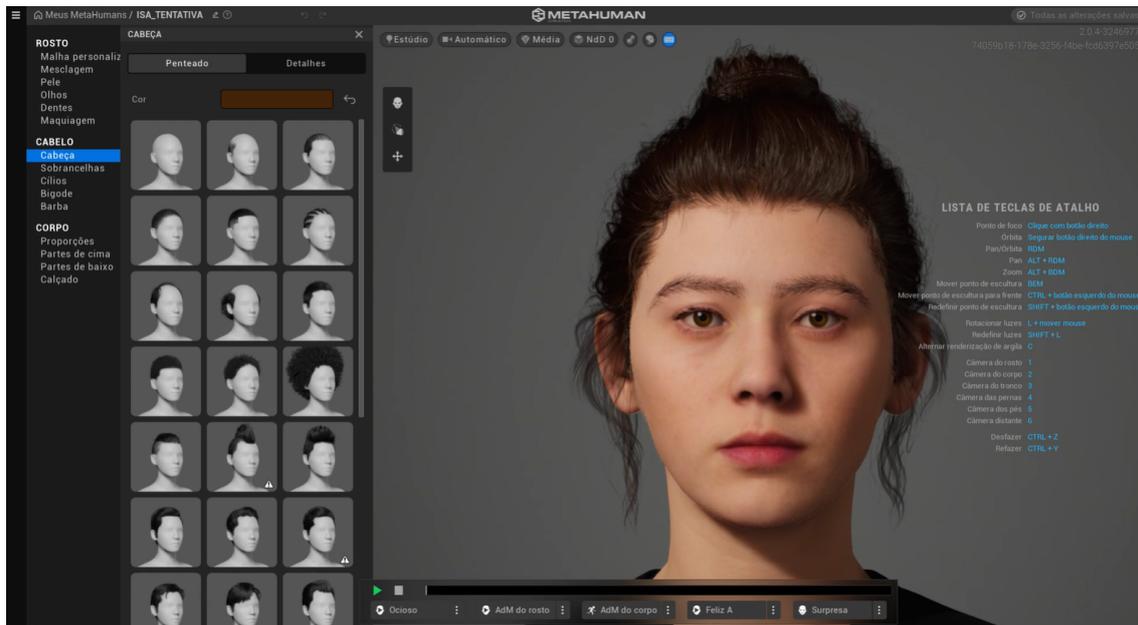


Figura 3. Na esquerda são os possíveis áreas de customização e tipos de cabelos e na direita o estado atual em tempo real do modelo do *MetaHuman*.

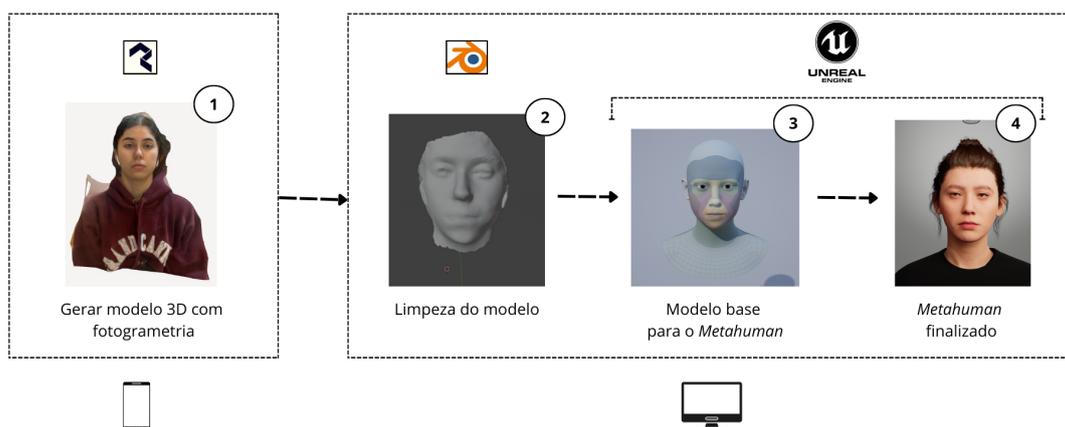


Figura 4. Processo de criação de um *MetaHuman* funcional. 1) Utiliza-se a técnica de fotogrametria para capturar imagens detalhadas do rosto do indivíduo. 2) Realiza uma limpeza do modelo 3D no Blender. 3) O modelo 3D refinado é importado para a *Unreal Engine 5*, onde é utilizado para gerar o modelo base do *MetaHuman*. 4) *MetaHuman* finalizado



Figura 5. Imagem da tela do aplicativo *Live Link Face*

A próxima etapa do processo é a consolidação das expressões do humano real com o *MetaHuman* criado. Nesta etapa do processo será utilizado um aplicativo móvel proprietário da Epic Games chamado *Live Link Face*⁸, o qual permite capturar as expressões e características da face, utilizando as ferramentas do *iPhone*, como a câmera *TrueDepth* que captura dados do rosto com precisão, analisando e projetando milhares de pontos invisíveis para criar um mapa de profundidade do rosto, além de capturar uma imagem infravermelha da face como pode ser observado na Figura 5.

Nesta etapa, é necessário que sejam realizadas movimentações do rosto para o ajuste posterior na *Unreal Engine 5*, como pode ser observado nas etapas 1 e 3 da Figura 7. Primeiro, são realizadas movimentações básicas da cabeça (frontal, esquerda e direita) para garantir que o modelo 3D possa ser ajustado com precisão na *Unreal Engine 5*. A seguir, podem ser realizadas quaisquer expressões e movimentos da face, pois estes serão utilizados para a reprodução posterior do *MetaHuman*, como pode ser observado nas etapas 2 e 4 da Figura 7.

Como o foco deste estudo é realizar uma análise sobre as 6 emoções descritas por [Barrett et al. 2019], após a modelagem da face, estas serão os movimentos e expressões executados pelos 3 indivíduos.

Na figura 6 são apresentados os três indivíduos, cada um ao lado de sua contraparte virtual criada usando a tecnologia *MetaHuman*. O grupo consiste em um homem e duas mulheres, ilustrando tanto suas imagens reais quanto as representações virtuais.

⁸<https://dev.epicgames.com/community/learning/tutorials/lEYe/unreal-engine-facial-capture-with-live-link>

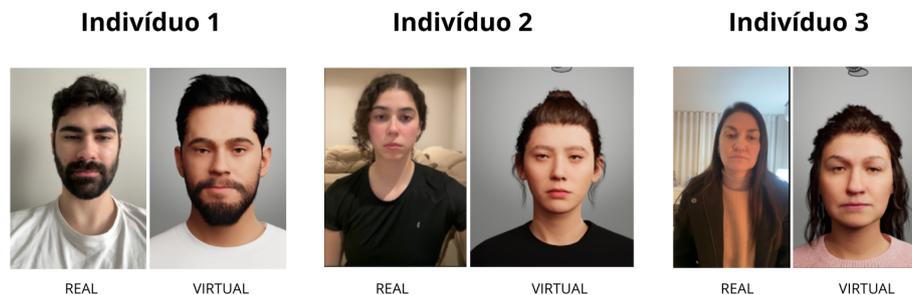


Figura 6. Três indivíduos, cada um ao lado de sua contraparte virtual criada usando a tecnologia *MetaHuman*, onde o indivíduo 1 é masculino enquanto o 2 e 3 são femininos.

5.2. Avaliação do nível de intensidade das expressões

Para analisar algo subjetivo como expressões faciais de emoções humanas, utilizamos o conceito de *Action Units* (AUs) proposto por Paul Ekman [Paul Ekman 1978], em conjunto com seis emoções básicas: Raiva, Medo, Nojo, Felicidade, Tristeza e Surpresa [Barrett et al. 2019]. Com base nesses conceitos, quantificamos a intensidade das AUs ativadas por essas expressões. A análise da intensidade foi realizada tanto para a face real quanto para a virtual dos três indivíduos. Cada indivíduo teve suas seis expressões faciais analisadas, as quais foram todas posadas e não espontâneas.

Para gerar os valores de intensidade, utilizamos o programa OpenFace⁹. Ele usa técnicas avançadas de visão computacional e aprendizado de máquina para detectar pontos-chave no rosto e determinar quais AUs estão presentes e com quais intensidade. Na Tabela 3, podemos observar quais AUs são ativadas com mais intensidade para cada expressão emocional, de acordo com a literatura [Mavadati et al. 2013] e [Ghayoumi and Bansal 2016]. Com base nessas informações, as Tabelas 4, 5, 6, 7, 8 e 9 apresentam o nível da intensidade de cada *Action Unit* levando em consideração os três indivíduos estudados. A Seção 6 apresenta e discute os resultados obtidos.

6. Análise dos Resultados

Nesta seção, serão analisados os valores de intensidade das *Action Units* para cada expressão exibida pelos indivíduos, conforme ilustrado nas Figuras 8 a 13. Estas figuras apresentam a soma das intensidades de todas as AUs ativadas para cada expressão de referência. As análises incluíram comparações entre a suavização das expressões faciais dos indivíduos nas representações real e virtual. Adicionalmente,

⁹<https://github.com/TadasBaltrusaitis/OpenFace>

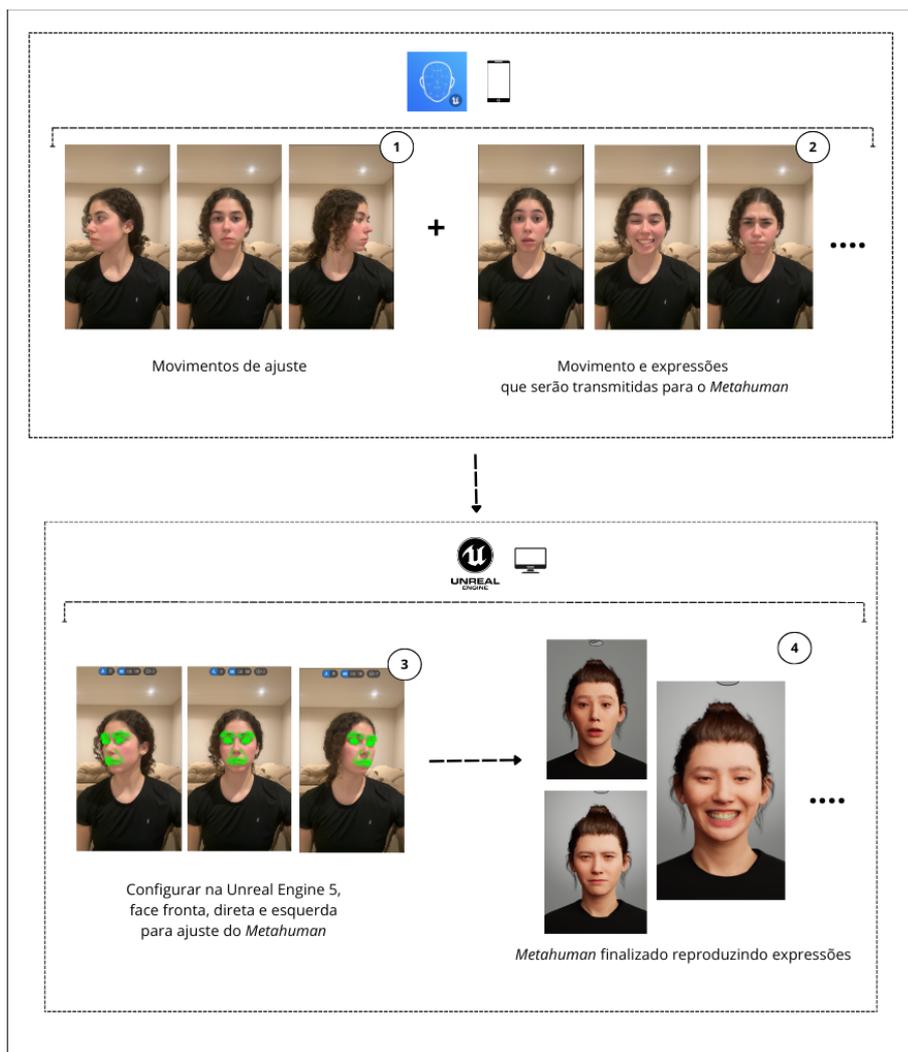


Figura 7. Processo da transferência das expressões do humano real para o virtual. 1) Realiza movimentos de ajuste do Live link face. 2) Realiza quaisquer movimentos e expressões. 3) Ajusta a imagens frontais e laterais para o MetaHuman. 4) MetaHuman reproduzindo as expressões.

Emoção	Action Units (AU) Presentes
Raiva	4, 5, 7, 10, 17, 22, 23, 24, 25, 26
Medo	1, 2, 4, 5, 20, 25, 26
Nojo	9, 10, 16, 17, 25, 26
Felicidade	6, 12
Tristeza	1, 4, 11, 15, 17
Surpresa	1, 2, 5, 26

Tabela 3. Action Units (AU) presentes para cada emoção [Mavadati et al. 2013] [Ghayoumi and Bansal 2016]

Name	AU06	AU12
CG_individuo_1_masculino	1.15	2.72
CG_individuo_2_feminino	1.58	2.65
CG_individuo_3_feminino	1.14	2.7
REAL_individuo_1_masculino	2.62	3.01
REAL_individuo_2_feminino	1.96	2.82
REAL_individuo_3_feminino	1.4	3.07

Tabela 4. Valores das AUs 06 e 12 para a emoção de Felicidade

serão feitas comparações com outras ferramentas de conversão de humanos reais para avatares virtuais para a expressão da emoção de felicidade, conforme discutido por [Vitor Miguel Xavier Peres and Musse. 2023].

6.1. Análise dos níveis de suavização

Com base na Tabela 12, que apresenta as análises de intensidade emocional por indivíduo e por emoção, foi realizada uma análise detalhada sobre os níveis de suavização das expressões faciais entre indivíduos reais e seus correspondentes virtuais.

A suavização foi calculada como a porcentagem de redução na intensidade das expressões faciais ao serem convertidas de rostos reais para rostos virtuais. A seguir discutimos algumas observações que puderam se realizadas.

A suavização da expressão de felicidade aparece em todos os indivíduos, com valores variando de aproximadamente 11% a 31%. Isso sugere que, embora a suavização seja perceptível em todos, ela pode variar dependendo do indivíduo.

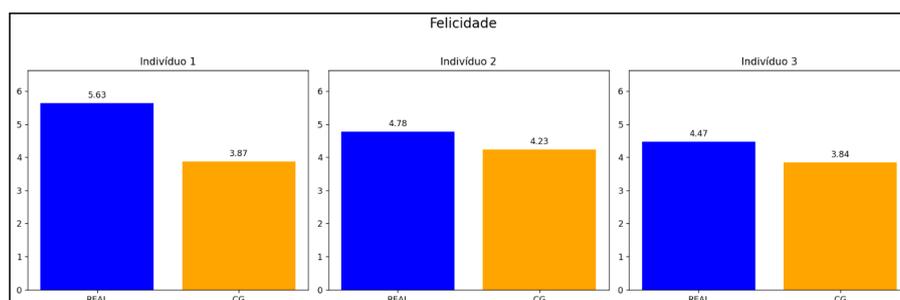


Figura 8. Soma total dos valores de intensidades das Action Units ativadas para os indivíduos 1, 2 e 3 tanto real quanto virtual(CG) para a expressão de felicidade.

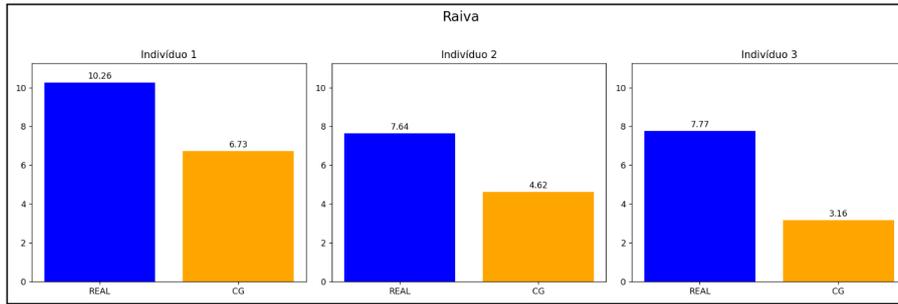


Figura 9. Soma total dos valores de intensidades das *Action Units* ativadas para os indivíduos 1, 2 e 3 tanto real quanto virtual(CG) para a expressão de raiva.

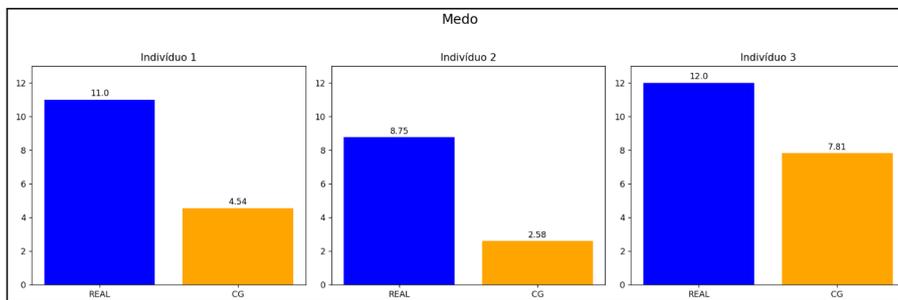


Figura 10. Soma total dos valores de intensidades das *Action Units* ativadas para os indivíduos 1, 2 e 3 tanto real quanto virtual(CG) para a expressão de medo.

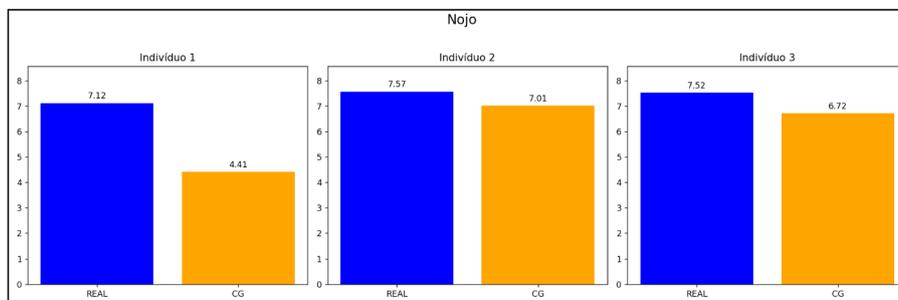


Figura 11. Soma total dos valores de intensidades das *Action Units* ativadas para os indivíduos 1, 2 e 3 tanto real quanto virtual(CG) para a expressão de nojo.

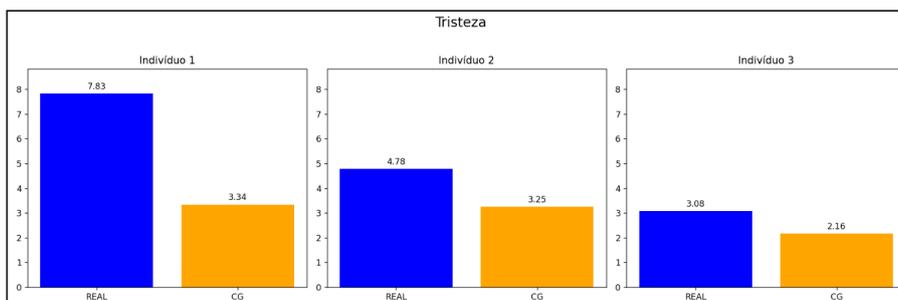


Figura 12. Soma total dos valores de intensidades das *Action Units* ativadas para os indivíduos 1, 2 e 3 tanto real quanto virtual(CG) para a expressão de tristeza.

Name	AU04	AU05	AU07	AU10	AU17	AU23	AU25	AU26
CG_individuo_1_masculino	1.77	0.71	0.4	1.17	1.94	0.74	0.0	0.0
CG_individuo_2_feminino	1.69	0.0	0.57	1.27	0.6	0.49	0.0	0.0
CG_individuo_3_feminino	1.16	0.0	1.18	0.0	0.0	0.82	0.0	0.0
REAL_individuo_1_masculino	4.83	0.05	2.38	1.03	1.72	0.0	0.25	0.0
REAL_individuo_2_feminino	3.56	0.0	2.01	0.41	1.2	0.46	0.0	0.0
REAL_individuo_3_feminino	3.46	0.15	3.58	0.09	0.49	0.0	0.0	0.0

Tabela 5. Valores das AUs 04, 05, 07, 10, 17, 23, 25 e 26 para a emoção de Raiva

Name	AU01	AU02	AU04	AU05	AU20	AU25	AU26
CG_individuo_1_masculino	1.42	2.14	0.0	0.98	0.0	0.0	0.0
CG_individuo_2_feminino	0.09	0.0	0.83	1.06	0.6	0.0	0.0
CG_individuo_3_feminino	0.82	1.81	1.28	0.0	0.6	1.67	1.63
REAL_individuo_1_masculino	2.06	3.68	0.82	3.24	0.69	0.51	0.0
REAL_individuo_2_feminino	2.01	1.7	1.58	2.96	0.24	0.02	0.24
REAL_individuo_3_feminino	2.49	4.37	1.03	1.12	0.0	1.79	1.2

Tabela 6. Valores das AUs 01, 02, 04, 05, 20, 25 e 26 para a emoção de Medo

A raiva apresenta uma das maiores taxas de suavização, especialmente para o Indivíduo 3, onde a suavização chega a quase 60%. Isso pode indicar que a conversão de expressões de raiva para CG perde mais da intensidade original comparada a outras emoções.

O medo também mostra uma alta taxa de suavização, especialmente para o Indivíduo 2, onde atinge mais de 70%. A variação entre indivíduos é significativa nesta emoção, sugerindo que a intensidade do medo é particularmente difícil de ser transferida do real para o virtual.

Para a emoção de nojo, a suavização varia amplamente, com uma redução mínima para o Indivíduo 2 (7.40%) e maior para o Indivíduo 1 (38.07%). Isso sugere que a expressão de nojo pode ser melhor preservada em alguns indivíduos.

A tristeza também apresenta uma suavização significativa, especialmente para o Indivíduo 1, onde a redução é de mais de 57%. As variações entre indivíduos sugerem que a intensidade da tristeza também pode ser difícil de preservar na conversão para virtual.

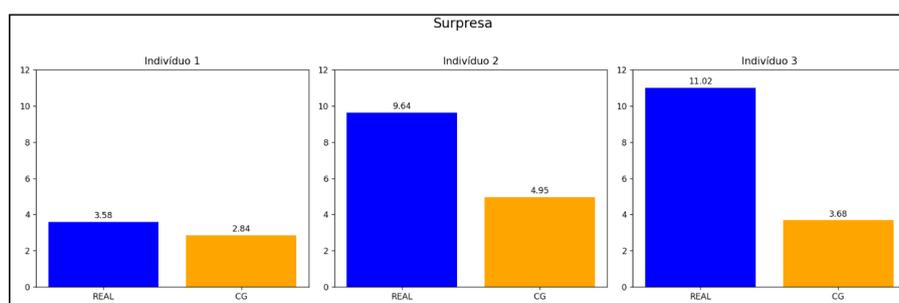


Figura 13. Soma total dos valores de intensidades das Action Units ativadas para os indivíduos 1, 2 e 3 tanto real quanto virtual(CG) para a expressão de surpresa.

Name	AU09	AU10	AU17	AU25	AU26
CG_individuo_1_masculino	1.55	0.42	0.34	2.1	0.0
CG_individuo_2_feminino	2.05	2.05	0.15	2.76	0.0
CG_individuo_3_feminino	1.71	2.8	0.0	2.21	0.0
REAL_individuo_1_masculino	3.97	0.69	0.02	2.44	0.0
REAL_individuo_2_feminino	1.73	1.89	0.36	3.44	0.15
REAL_individuo_3_feminino	1.91	2.89	0.0	2.72	0.0

Tabela 7. Valores das AUs 09, 10, 17, 25 e 26 para a emoção de Nojo

Name	AU01	AU04	AU15	AU17
CG_individuo_1_masculino	0.02	0.32	1.06	1.94
CG_individuo_2_feminino	0.32	1.97	0.66	0.3
CG_individuo_3_feminino	0.23	1.04	0.68	0.21
REAL_individuo_1_masculino	1.38	3.3	1.14	2.01
REAL_individuo_2_feminino	1.93	2.19	0.66	0.0
REAL_individuo_3_feminino	1.48	0.52	0.59	0.49

Tabela 8. Valores das AUs 01, 04, 15 e 17 para a emoção de Tristeza

A surpresa apresenta uma ampla gama de suavizações, com a maior redução observada no Indivíduo 3 (66.61%). A variação sugere que a intensidade da surpresa pode ser significativamente afetada pela conversão para virtual.

Os resultados de suavização das expressões observados no presente trabalho confirmam as observações de [Mäkäräinen et al. 2014]. Os autores sugeriram que, ao transferir uma expressão facial diretamente de um domínio real para um domínio virtual, a expressão tende a parecer menos intensa devido à natureza mais abstrata dos rostos virtuais. Os presentes resultados da Tabela 12 estão alinhados com essa hipótese, demonstrando uma redução perceptível nos níveis de intensidade das expressões faciais quando convertidas para rostos virtuais. Esses resultados indicam que a suavização das intensidades é uma característica consistente da transição de expressões faciais de humanos reais para representações computacionais independente da emoção e expressão representada na face real.

As análises realizadas, conforme representado nas Tabelas 10 e 11, fornecem resultados sobre a variação na suavização das expressões faciais ao serem transferidas de uma representação real para uma virtual. A Tabela 10 destaca diferenças significativas entre as emoções em termos de como elas são afetadas pelo processo de digitalização.

A suavização das expressões faciais revela uma tendência de perda maior de intensidade em emoções mais complexas (mais individualizadas e podem ativar mais músculos) e intensas, como o medo e a raiva, que registram médias de suavização de 54.72% e 44.43%, respectivamente. Esses resultados indicam que essas emoções, que frequentemente envolvem uma gama mais ampla de músculos faciais, são mais suscetíveis a serem diminuídas em intensidade quando modeladas digitalmente. O desvio padrão elevado, especialmente no medo (19.16%) e na surpresa (22.78%), sugere uma grande variação na qualidade da transposição dessas emoções para o ambiente virtual, refletindo a complexidade em capturar e reproduzir fielmente suas nuances dinâmicas.

Name	AU01	AU02	AU05	AU26
CG_individuo_1_masculino	0.73	1.26	0.85	0.0
CG_individuo_2_feminino	1.68	2.05	0.43	0.79
CG_individuo_3_feminino	0.89	1.77	0.0	1.02
REAL_individuo_1_masculino	0.48	2.18	0.92	0.0
REAL_individuo_2_feminino	2.99	3.55	2.39	0.71
REAL_individuo_3_feminino	2.31	4.1	4.12	0.49

Tabela 9. Valores das AUs 01, 02, 05 e 26 para a emoção de Surpresa

Por outro lado, emoções como felicidade e nojo apresentam uma média de suavização mais baixa, de 18.96% e 18.70%, respectivamente, com um desvio padrão que indica uma menor variação entre as transições, sugerindo que estas são relativamente mais fáceis de serem preservadas em sua intensidade durante o processo de conversão para CG.

A análise detalhada da Tabela 11 revela detalhes sobre o impacto sobre cada indivíduo separadamente. Foi observado que para todos os três indivíduos, há uma redução consistente na intensidade total das expressões quando transformadas de real para CG. Por exemplo, o Indivíduo 1 apresenta uma soma de intensidades de 45.42 para expressões reais e uma redução para 25.74 em sua versão CG, destacando uma significativa perda de detalhes emocionais na conversão. Essa tendência é consistente nos outros indivíduos também, com o Indivíduo 2 mostrando uma soma de 43.14 no real e 26.64 no CG, e o Indivíduo 3 de 45.84 no real para 27.30 no CG.

Emoção	Média de Suavização (%)	Desvio Padrão de Suavização (%)
Felicidade	18.96	10.27
Raiva	44.43	12.73
Medo	54.72	19.16
Nojo	18.70	15.30
Tristeza	39.73	15.13
Surpresa	45.28	22.78

Tabela 10. Média e desvio padrão da porcentagem de suavização das expressões faciais para cada emoção

Indivíduo	Tipo	Média	Desvio Padrão	Soma das Intensidades
Indivíduo 1	Real	7.57	2.63	45.42
Indivíduo 1	CG	4.29	1.50	25.74
Indivíduo 2	Real	7.19	1.91	43.14
Indivíduo 2	CG	4.44	1.99	26.64
Indivíduo 3	Real	7.64	3.36	45.84
Indivíduo 3	CG	4.55	2.15	27.30

Tabela 11. Comparação das médias, desvios padrão e soma das intensidades das emoções reais e CG para cada indivíduo.

Nome da Emoção	Indivíduo	Real	CG	Porcentagem de Suavização
Felicidade	Indivíduo 1	5.63	3.87	31.26%
Felicidade	Indivíduo 2	4.78	4.23	11.55%
Felicidade	Indivíduo 3	4.47	3.84	14.08%
Raiva	Indivíduo 1	10.26	6.73	34.42%
Raiva	Indivíduo 2	7.64	4.62	39.53%
Raiva	Indivíduo 3	7.77	3.16	59.34%
Medo	Indivíduo 1	11.00	4.54	58.73%
Medo	Indivíduo 2	8.75	2.58	70.51%
Medo	Indivíduo 3	12.00	7.81	34.92%
Nojo	Indivíduo 1	7.12	4.41	38.07%
Nojo	Indivíduo 2	7.57	7.01	7.40%
Nojo	Indivíduo 3	7.52	6.72	10.64%
Tristeza	Indivíduo 1	7.83	3.34	57.34%
Tristeza	Indivíduo 2	4.78	3.25	31.97%
Tristeza	Indivíduo 3	3.08	2.16	29.87%
Surpresa	Indivíduo 1	3.58	2.84	20.67%
Surpresa	Indivíduo 2	9.64	4.95	48.55%
Surpresa	Indivíduo 3	11.02	3.68	66.61%

Tabela 12. Comparação das intensidades das expressões faciais reais e CG e a porcentagem de suavização das expressões

6.2. Comparação com a literatura: Análise da transferência de expressão de felicidade

Ao analisar a suavização das intensidades das AUs para a emoção de Felicidade em indivíduos reais e suas contrapartes virtuais, foi observada uma suavização consideravelmente menor em comparação com os dados apresentados no estudo de [Vitor Miguel Xavier Peres and Musse. 2023], como é possível observar na Tabela 10 com a média da porcentagem de suavização para cada indivíduo para a emoção felicidade. Neste artigo, é mencionado que a suavização para AU6 e AU12 em rostos posados alcançou uma redução média de aproximadamente 70% e 32%, respectivamente, para ambos os gêneros. Em contraste, em nosso estudo, os dados revelam uma suavização de 31% para a AU6 e apenas 10% para a AU12, quando observamos o indivíduo com maior redução. Ainda, a felicidade em média obteve a menor média de suavização, de acordo com a Tabela 10.

Essa diferença nos resultados pode ser atribuída a vários fatores, incluindo as metodologias de captura de expressões faciais e as tecnologias de renderização usadas para criar os rostos virtuais, como as utilizadas no estudo [Vitor Miguel Xavier Peres and Musse. 2023], Emoca[Danecek et al. 2022], Deca[Feng et al. 2021] e Deep3D[Deng et al. 2019]. Particularmente, a utilização da tecnologia *MetaHuman* no presente estudo pode ter desempenhado um papel crucial. O *MetaHuman* é conhecido por sua capacidade de gerar avatares digitais altamente realistas com uma fidelidade de expressão facial notavelmente alta, sendo considerado o estado-da-arte na área. Isso sugere que as ferramentas e técnicas empregadas foram capazes de manter uma maior fidelidade às expressões originais, possivelmente devido à

sofisticação avançada da tecnologia *MetaHuman*.

Além disso, a discussão de [Vitor Miguel Xavier Peres and Musse. 2023] ressalta que a natureza abstrata dos rostos CG pode contribuir para a percepção de uma intensidade reduzida em expressões diretamente copiadas de um domínio para outro. Essa observação é consistente com nossos resultados, embora a magnitude da suavização seja significativamente menor em nosso estudo. Tal diferença nos resultados enfatiza a variabilidade potencial na representação de expressões faciais em ambientes virtuais, que pode variar amplamente dependendo das especificidades do software e das técnicas de modelagem empregadas, além da própria expressão do ator/atriz. A capacidade do *MetaHuman* de minimizar essa perda de intensidade é um indicativo do potencial da tecnologia para melhorar a forma como as expressões faciais são produzidas e percebidas em contextos digitais.

7. Considerações finais

Este trabalho apresentou um estudo aprofundado sobre a viabilidade e eficácia de transferir expressões faciais de indivíduos reais para avatares virtuais, utilizando tecnologias recentes como a *Unreal Engine 5* e a plataforma *MetaHuman*. Através da metodologia proposta, foram examinados os detalhes da reconstrução facial e a transferência de expressões emocionais, fundamentadas no conceito de *Action Units (AUs)* proposto por [Paul Ekman 1978], associadas a diferentes emoções humanas, como felicidade, raiva, medo, nojo, tristeza e surpresa.

A utilização da tecnologia *MetaHuman* demonstrou ser fundamental na obtenção de modelos 3D altamente realistas e expressivos. O uso combinado de fotogrametria e ferramentas avançadas de captura e modelagem permitiu uma representação detalhada das características faciais dos indivíduos, proporcionando uma base sólida para a análise subsequente das expressões faciais.

O estudo revelou que, embora a suavização das expressões ao serem transferidas para um domínio virtual seja inevitável, as técnicas empregadas neste trabalho permitiram minimizar essas distorções. A comparação dos níveis de suavização das AUs com estudos anteriores, como os de [Vitor Miguel Xavier Peres and Musse. 2023], destacou a eficácia das abordagens utilizadas, demonstrando uma menor perda de intensidade das expressões faciais nos avatares *MetaHuman* em comparação com outros métodos.

Este trabalho contribuiu para o campo da computação gráfica e da animação facial, oferecendo informações e análises sobre as limitações e possibilidades das tecnologias atuais. A capacidade de reproduzir e investigar a acurácia das expressões faciais humanas em avatares digitais com alta fidelidade abre novas possibilidades para a indústria de jogos, filmes e realidade virtual, onde a representação precisa de emoções humanas é crucial.

Em suma, os resultados obtidos reforçam a viabilidade de usar avatares *MetaHuman* para reproduzir expressões humanas com precisão e detalhe, contribuindo para o avanço das tecnologias de animação facial e destacando o potencial crescente da *Unreal Engine 5* como uma ferramenta essencial no desenvolvimento de conteúdo digital humano realista.

Para trabalhos futuros, seria ideal expandir o escopo da pesquisa para incluir uma

variedade maior e mais diversificada de participantes. Isso incluiria pessoas de diferentes aparências, idades e gêneros. Ampliar a base de dados desta forma não só aumentaria a robustez dos resultados, como também permitiria testar a eficácia das tecnologias *MetaHuman* e *Unreal Engine 5* em um espectro mais amplo de características faciais e expressões emocionais, que podem variar significativamente entre diferentes grupos demográficos. Ainda, isto permitiria desenvolver e implementar métodos de avaliação quantitativa mais sofisticados para medir a fidelidade das expressões faciais transferidas para os avatares digitais de forma mais detalhada.

8. Agradecimentos

Agradecimento a Vitor Miguel Xavier, doutorado do PPGCC-PUCRS, pelo auxílio no desenvolvimento do trabalho.

Referências

- Technical specifications for metahumans: Platform and lod specifications. <https://dev.epicgames.com/documentation/es-mx/metahuman/technical-specifications-for-metahumans/platform-and-lod-specifications>. Accessed: 25-11-2023.
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1):1–68.
- Danecek, R., Black, M. J., and Bolkart, T. (2022). EMOCA: Emotion driven monocular face capture and animation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20311–20322.
- Deng, Y., Yang, J., Xu, S., Chen, D., Jia, Y., and Tong, X. (2019). Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set. In *IEEE Computer Vision and Pattern Recognition Workshops*.
- Feng, Y., Feng, H., Black, M. J., and Bolkart, T. (2021). Learning an animatable detailed 3D face model from in-the-wild images. *Volume 40*.
- Ghayoumi, M. and Bansal, A. (2016). Unifying geometric features and facial action units for improved performance of facial expression analysis. *arXiv preprint arXiv:1606.00000*.
- Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., and Cohn, J. F. (2013). Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4(2):151–160.
- Mäkäräinen, M., Kätsyri, J., and Takala, T. (2014). Exaggerating facial expressions: A way to intensify emotion or a way to the uncanny valley? *Cognitive Computation*, 6:708–721.
- Paul Ekman, W. V. F. (1978). Facial action coding system: A technique for the measurement of facial movement.
- Vitor Miguel Xavier Peres, G. P. D. M. and Musse., S. R. (2023). Can we truly transfer an actor’s genuine happiness to avatars? an investigation into virtual, real, posed and spontaneous faces. *Symposium on Games and Digital Entertainment (SBGames 2023)*, (1):1–10.