

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



Bernardo Gomes Alevato

Do metaverso à inteligência artificial
Uma trajetória da percepção tridimensional

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Design da PUC-Rio como requisito parcial para obtenção do grau de Doutor em Design.

Jorge Lopes

Orientador

Luiz Velho

Coorientador

Rio de Janeiro,
abril de 2026



Bernardo Gomes Alevato

Do metaverso à inteligência artificial

Uma trajetória da percepção tridimensional

Tese de Doutorado

Tese de Doutorado apresentada como requisito parcial para a obtenção do grau de Doutor pelo Programa de Pós-graduação em Design da PUC-Rio.

Prof. Jorge Roberto Lopes

Orientador

Departamento de Artes e Design – PUC-Rio

Prof. Luiz Velho

Co-Orientador

Instituto Nacional de Matemática Pura e Aplicada – IMPA

Prof. Luiza Novaes

Departamento de Artes e Design – PUC-Rio

Prof. João de Sá Bonelli

Departamento de Artes e Design – PUC-Rio

Prof. Sergio Alex Kugland de Azevedo

Universidade Federal do Rio de Janeiro – UFRJ

Prof. Sergio Azevedo

Departamento de Artes e Design – PUC-Rio

Rio de Janeiro, 15 de abril de 2026

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Bernardo Gomes Alevato

Graduou-se em Comunicação Visual na PUC-Rio (Pontifícia Universidade Católica do Rio de Janeiro) em 1997. Atuou *como designer* nas áreas de *design* impresso e mídia digital desenvolvendo projetos de identidade visual, websites e aberturas de filmes de curta e longa-metragem. Fez mestrado na PUC-Rio, com tema abrangendo design em movimento e realidade virtual. Participa como professor de disciplinas do Departamento de Artes e Design da PUC-Rio.

Agradecimentos

“O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001”

“This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001”

À CAPES e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

À Sergio Azevedo, Pedro Luiz Von Seehausen, Vinicius Arcoverde, Gerson Ribeiro, Mariana Duarte, Carolina Navarro, Thaisa Coelho, Orlando Grillo, Gabriel Cardoso, Andrea Lennhof, Matteo Moriconi, Leonardo Massoto, Hamsa Wood, Heron Werner. E em especial à Luiz Velho e Jorge Lopes por acompanharem essa trajetória terem aberto tantos caminhos.

Ao VISGRAF Lab e ao IMPA: Isabella Costa, Fabio Suim, Daniel Perazzo, Mohara Nascimento. João Bonelli e ao LIFE, Luiza Novaes e ao EAI e aos colegas do dAD. Ao Lis, Eliane Garcia Juliana Gonçalves e Antonio Cabral Rebello e toda equipe. A Tauane Luzes

A Taianne Lobo, Fabiano Nunes e Helena Bevilaqua por tornarem este estudo mais poético. Sempre grato a André Ramos, Eloise Porto Ferreira, João Vitor da Silva e Claudia Bolshaw por me ajudarem em meu caminho.

À Clarice, com quem aprendi a ser pai. E a Melanie, por me dar uma família A Jurema e Sergio Alevato. Aos professores que participaram da Comissão Examinadora no exame de qualificação e na defesa da dissertação.

Resumo

Alevato, Bernardo Gomes; Lopes, Jorge. Um ambiente para integração de dados bibliográficos baseado em mediadores. Rio de Janeiro, 2026. 141p. Tese de Doutorado - Departamento de Artes e Design, Pontifícia Universidade Católica do Rio de Janeiro

O documento apresenta uma pesquisa em design voltada à construção de imagens tridimensionais para experiências em Realidade Expandida (XR), entendida no continuum realidade–virtualidade, com ênfase em utilização de Gaussian Splats e inteligência artificial generativa, observamos a convicção perceptiva e a sensação de presença. Partindo de uma trajetória de prática experimental e reflexão o texto organiza um repertório de casos e testes que tratam da representação de objetos, ambientes e humanos em uma abordagem sobre imagens de interesse histórico-cultural.

Estão descritos fluxos de construção de 3DGS, relações com a captura fotográfica/vídeo com sobreposição e paralaxe, possibilidades de edição manual. Acompanhamos o início da utilização de inteligência artificial generativa com experimentos de imagem, vídeo, 3D e treinamento adicional.

Elencamos técnicas e reflexões relativas ao formato de imagem e ao design de experiências imersivas somando organicidade ao amplo repertório que reside na modelagem por malhas.

Palavras-chave

Gaussian Splatting, Inteligência Artificial, 3D, Experiências imersivas e Realidades Expandidas, Realidade Virtual e Aumentada

Abstract

The document presents research in design focused on the construction of three-dimensional images for Augmented Reality (XR) experiences, understood within the reality-virtuality continuum, with an emphasis on the use of Gaussian Splats and generative artificial intelligence. We observe perceptual conviction and the sensation of presence. Starting from a trajectory of experimental practice and reflection, the text organizes a repertoire of cases and tests that deal with the representation of objects, environments, and humans in an approach to images of historical-cultural interest.

We describe 3DGS construction flows, relationships with photographic/video capture with overlay and parallax, and possibilities for manual editing. We follow the beginning of the use of generative artificial intelligence with experiments in image, video, 3D, and additional training.

We list techniques and reflections related to image format and the design of immersive experiences, adding organicity to the broad repertoire that resides in mesh modeling.

Keywords

Gaussian Splatting, Artificial Intelligence, 3D, Immersive Experiences and Augmented Realities, Virtual and Augmented Reality

Sumário

Agradecimentos	4
Resumo	5
Sumário	7
Lista de Figuras	9
Lista de abreviaturas	16
1 Introdução	1
2. Luz, cor e espaço	11
2.1. Captura e treinamento	15
2.2. Edição	19
2.3. Uma breve comparação e características	21
2.3. Aspectos visuais	26
3. Entre o intrínseco e o extrínseco	32
3.1 A câmera intrínseca	33
3.2 A câmera extrínseca	43
3.3 Distribuição	54
4. IA como ferramenta para expandir a realidade	61

4.1 Conectando e abrindo caixas	61
4.2 Preenchendo lacunas	72
4.3 Experimentos generativos 2D, treinamento e movimento	74
4.4 Objetos generativos 3D	82
4.5 Profundidade e significado / visão computacional	86
4.6 Vertendo o eixo temporal para profundidade	91
5. Representação humana e movimento	96
5.1 Ocean Dance	99
5.2 Mocap IA	102
5.3 Avatares	105
5.4 Luz, espaço, pessoas e movimento	109
6. Considerações Finais	113
7. Referências Bibliográficas	119

Lista de Figuras

Figura 1 - Grafismo criado para o grupo de estudos pensado como responsivo, isto é que pudesse ser flexível para aplicações em diferentes situações fonte: próprio autor	3
Figura 2 - Fotoescultura de François Willème. fonte:(Laric, [S.d.])	11
Figura 3 -Alfred Silvester, Pequena Nell. Em – 'A Loja de Antiguidades' de Charles Dickens, década de 1870-1880. Duas impressões em albúmen em formato estereoscópico. Fonte: Coleção de Artes Gráficas GAX	12
Figura 4 - Lenticular or 'Brewster' stereoscope, onde é possível notar a semelhança com headsets XR Fonte: ("180 years of 3D Royal Society", 2018)	13
Figura 5 - Gráfico do fluxo de treinamento básico de 3DGS	15
Figura 6 - Structure from motion- Mecanismo que observa pontos de interseção na imagem e posiciona em um espaço tridimensional, criando uma nuvem de pontos e o mapeamento dos pontos de vista na extração de colmap.	17
Figura 7 - breve sequência do treinamento em 3dgs. Disponível em	18
Figura 8 - Seleção realizada no Supersplat em 3DGS, separando a pessoa representada, João Vitor Silva, do ambiente fonte:	20
Figura 9 - Estrutura de modelo tridimensional / em baixo número de polígonos para deixar a estrutura visível Fonte: Próprio autor	23

Figura 10 - Estrutura de modelo em 3DGS / em baixo número de splats para deixar a estrutura visível Fonte: Próprio autor	24
Figura 11 – Uma forma dependente do ponto de vista Fonte: https://superspl.at/view?id=29f9b77d	26
Figura 12 - Imagem em 3DGS baseada em filme de 1906 de Oskar Messter que usa modelos-vivos para criar cenas de Adão e Eva (existe uma maçã na mão da modelo, de difícil visualização). Fonte	27
Figura 13 - Os elipsoides formam massas de cor através da popularização da nuvem de pontos	28
Figura 14 – Imagem do autor - Fotorrealismo é facilitado através da técnica 3DGS, disponível em: https://superspl.at/scene/c7f14114	28
Figura 15 - Edmund Fraser Gaussian Splat	29
Figura 16 - Dani Bittel 3DGS	30
Figura 17 - Imagem do autor apresenta a imagem incompleta, com splats visíveis como parte da linguagem	31
Figura 18 - Em uma vista superior a disposição de navegação intrínseca e extrínseca	32
Figura 19 - Sala LAPID no espaço XR-Showcase dentro da plataforma Spatial.io, disponível em: https://www.spatial.io/s/XR-Showcase-614666928c2af60001d23190?share=2973599184567383091	34
Figura 20 - processo de <u>tratamento</u> e textura do modelo do Fóssil da Luzia, da esquerda para direita o modelo bruto com milhões de polígonos, o segundo gerado através de remesh do zbrush, o	

terceiro em retoque, abaixo à esquerda, e o quarto com textura aplicada no adobe substance painter. Disponível no endereço acima	36
Figura 21 - Um estudo de navegação proposta pelo cenário e pelos objetos dispostos organicamente pelo grupo	37
Figura 22 - Comparação entre modelos gerados a partir de tomografia e fotogrametria e fotos do artefato	38
Figura 23 - Modelo gerado em malha	39
Figura 24 - O conjunto de imagens do caixão de Sha Amun em Su foi utilizado para o treinamento, ao treinar um pouco da sala onde foram feitas as imagens fica visível.	40
Figura 25 - processo de criação do modelo para geração da máscara	40
Figura 26 - Modelo final em 3DGS, disponível para visualização em https://superspl.at/scene/7d8feca	41
Figura 27 - Imagem em 3DGS da cartonagem da múmia Irthorru, realizada através do conjunto de imagens de autoria de Daniel Pett, no British Museum, disponível em https://superspl.at/scene/1431aac0	42
Figura 28 - Modelo 3DGS sinético feito a partir do Blender utilizando shader de vidro e uma textura de ossos disponível em: https://superspl.at/scene/a0d75ae3	43
Figura 29 - O nicho com as esculturas de Neferhotep e sua esposa é a expressão iconográfica do sítio	44
Figura 30 - Professora María Violeta Pereyra na TT49	45
Figura 31 - Imagens geolocalizadas geradas pelo Google Earth Studio	46

Figura 32 - O trabalho de videografismo foi realizado para auxiliar a compreensão entre imagens e locução	47
Figura 33 - Vinheta de abertura com o design do título	48
Figura 34 - Metaverso criado a partir de fotogrametria	49
Figura 35 - Experimento de ambiente e navegação	50
Figura 36 - Sala do Centro Pi em 3DGS	51
Figura 37 - Experimento de captura de ambiente	52
Figura 38 - Experimento de captura de ambiente	53
Figura 39 - Captura e visualização através do Hyscape da Meta	54
Figura 40 - Exemplos de utilização em realidade aumentada	56
Figura 41 - Metaverso criado a partir do projeto V-Horus	57
Figura 42 - Interface de "caixas" ou nós do ComfyUI	63
Figura 43 – Exemplos de imagens utilizadas no treinamento	64
Figura 44 - Primeiros resultados da experiência	67
Figura 45 - A Resultados de gerações do prompt “Portrait of Woman”	67
Figura 46 - Figura 4: Imagens geradas com o modelo utilizando o prompt “man with hat”	68
Figura 47 - Figura 35 - Figura 5: Imagens geradas nos modelos Archer Diffusion, Arcane Diffusion e SynthwavePunk v2 utilizando o prompt “man with hat”	68
Figura 48 - Resultados utilizando LORA e SDXL	69

Figura 49 - Modelo de tomografia texturizado e retocado manualmente para o projeto V-Horus	73
Figura 50 - A imagem já com fundo recortado e sua interpretação através do Stable Diffusion, quadro por quadro	75
Figura 51 - Um sentido de agência na interpretação dos quadros	76
Figura 52 - Etapas da animação "CORES". Da esquerda para direita, a captura de movimento pura realizada no IMPA pelo grupo Space-XR. Um avatar criado no Blender através do complemento MBLab. O Terceiro um tratamento sobre o tempo realizado em after effects e o quarto a imagem gerada pelo Stable Diffusion.	77
Figura 53 - Exemplo de imagens que fizeram parte do treinamento adicional utilizado em Gestos	78
Figura 54 – A imagem original, o guia de controle de profundidade em cima, o guia de estrutura corporal embaixo e o resultado final. Os guias de controle para profundidade e estrutura corporal, MiDas e DWpose facilitam a reinterpretação da imagem.	79
Figura 55 - Imagens do filme criado inteiramente em IA	81
Figura 56 - fotos de Bjørn Christian Tørrissen	83
Figura 57 - Mostramos os modelos lado a lado em modo comparativo com: Um modelo retirado do Scan the City, Trellis 1, Trellis 2, na linha de baixo o Hitem3d e o Huyuan 3d 3 e o 2.	84
Figura 58 - Figuras estilizadas inspiradas na Venus de Willendorf	85
Figura 59 - Imagens geradas pelo modelo Depth Anything 2	87
Figura 60 - Imagens de profundidade e sua aplicação gráfica	88

Figura 61 - 3DGS criado a partir de pintura <i>Senhora sentada ao virginal</i> (1673) de Johannes Vermeer.	89
Figura 62 - O SAM detecta, segmenta e reconstrói objetos tridimensionais	90
Figura 63 - 3DGS baseado em Natureza Morta com Flores. da Mota, Agostinho. Fonte: Acervo Pinacoteca de São Paulo, disponível em: https://superspl.at/scene/2abdbccd	92
Figura 64 - Tela verde e rotação utilizando Kling 3.1	93
Figura 65 - Nuvem de pontos e câmeras criadas através do vídeo generativo	94
Figura 66 - Alguns ângulos do 3DGS, disponível em: https://superspl.at/scene/4a00963b	94
Figura 67 - Étienne-Jules Marey criou imagens cuja grafia inspirou a representação do mocap. Fonte: (Teston, 2004)	97
Figura 68 - Um dos primeiros estudos de movimento realizados por Muybridge para investigar o galope de cavalo. Fonte: (Q190568 et al., [S.d.])	98
Figura 69 - Espaço de experimentos em metaverso com modelo tridimensional com movimento importado do mixamo para o Blender 3d.	99
Figura 70 - Registro da captura de movimento de Thaisa Martins (com Luiz Velho em diálogo sobre a captura em curso), em sequência, Fabiano Nunes e Helena Bevilaqua.	100
Figura 71 - Imagem da experiência com performance de Helena Bevilaqua capturada no VISGRAF/IMPA.	101

Figura 72 - Exploração gráfica através de captura de movimento	102
Figura 73 - Movimento capturado por 3 cameras utilizando mimem.ai	103
Figura 74 - Utilização do modelo SAM 3D Body para avaliar imagem	104
Figura 75 - O sistema Kimodo funciona com prompts e posicionamento dos <i>rigs</i>	105
Figura 76 - Imagem gerada em 3DGS, disponível em:	109
Figura 77 - Um vídeo em IA foi utilizado como teste para gerar um 4DGS utilizando o Apple Sharp	110
Figura 78 - O 4DGS gerado através de frame de vídeo e IA	111

Lista de abreviaturas

3DGS – 3D Gaussian Splatting

3DoF – 3 Degrees of freedom

6DoF – 6 Degrees of freedom - 6 graus de liberdade, a soma entre 3 ângulos de movimento da cabeça (lateral, vertical e inclinação) e o deslocamento x, y, z.

AR – Augmented Reality (Realidade Aumentada)

CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

HCI – Human Computer Interaction (Interação humano-computador)

HMD – Head mounted display – Óculos de realidade expandida

MUDs –

SAM – *Segmentation Anything Meta*

XR – Expanded reality (Realidade expandida)

VR – Virtual Reality (Realidade virtual)

VRML – *Virtual Reality Modeling Language*

1.

Introdução

Podemos verificar que historicamente a humanidade constrói espaços de imersão, como podemos observar desde as paredes pintadas no paleolítico, passando por igrejas como a Capela Sistina, pinturas panorâmicas e, posteriormente, o cinema em 180 graus. Observamos esta prática até os dias de hoje, em modo físico e posteriormente, digital. São utilizados como recursos a perspectiva, a luz, a cor e a composição, distribuídas em espaços bi e tridimensionais (Grau, 2004). Nesta pesquisa, buscamos uma leitura específica orientada a aspectos da imagem tridimensional, processos construtivos em XR, entendido no continuum realidade–virtualidade (Milgram et al., 1994) como forma de apresentar ambientes, objetos e performances de interesse cultural.

A imagem em realidade virtual tem por característica ser navegável pelo olhar, sendo esse o primeiro item interativo. Em um vídeo 360 existe a escolha para onde olhar (giro da cabeça, 3DoF- 3 graus de liberdade), e em uma imagem gerada através de malhas é possível deslocar-se (6DoF- giro da cabeça+ deslocamento X,Y,Z). Deste modo, a percepção de presença do virtual ocorre quando a coerência entre movimentos do corpo retorna visualmente. Através do olhar são despertadas sensações físicas a partir da ilusão de tridimensionalidade, assim o formato e as características da imagem tridimensional em si são projetados para afetar a experiência em modo primordial (GRAU, 2003).

Esta pesquisa se insere no contexto de consolidação gradual de dispositivos e plataformas de **realidade virtual** e **realidade aumentada**. Considera-se que a linguagem da tridimensionalidade imersiva interativa está em construção, pois introduz propriedades narrativas e comunicacionais que dependem de presença, localização perceptiva e agência. Nas experiências apresentadas, o estado de *flow* (Csikszentmihalyi, 1990) em aprendizado veloz é desejável: quase que instantaneamente a pessoa sem

experiência prévia percebe como operar o sistema utilizado, com breves instruções. Por isso, são projetos sem dificuldade de uso, e rápido aprendizado de uso, assim o foco é a própria fruição da representação imersiva como parte principal da experiência. Desse modo, as imagens aqui trabalhadas estão disponíveis no endereço <http://www.alevato.xyz/x-imgs>, juntamente com uma edição digital.

Esta pesquisa iniciou-se em 2022 em plena pandemia de COVID 19, que isolou as pessoas, causou muitos danos e promoveu mudanças de hábitos. Nesse momento foi formado um grupo de pesquisa sobre ambientes imersivos chamado Space-XR.

Os tópicos abordados são extraídos de alguns projetos realizados pelo grupo de estudo em realidades expandidas Space-XR. O grupo Space-XR teve características interdisciplinares e foi composto por pesquisadores de 3 centros de pesquisa: VISGRAF Lab no IMPA, LAPID e SESHAT do Museu Nacional e BIODESIGN na PUC-Rio. O grupo composto por Luiz Velho, Jorge Lopes, Sergio Azevedo, Vinicius Arcoverde, Gerson Ribeiro, Mariana Duarte, Carolina Navarro, Thaisa Coelho, Orlando Grillo e Gabriel Cardoso, fez diversos experimentos usando realidade virtual, aumentada, captura de movimento, pesquisando sobre mídias, expressão artística, medicina, divulgação científica e construção de conhecimento.

Em vários capítulos desta pesquisa é utilizado como base de memória dos experimentos do grupo o livro Realidade Expandida, Novas Mídias e IA (Azevedo et al., 2025), e o website relacionado (“REALIDADE EXPANDIDA”, 2025)¹, utilizado principalmente para acesso direto aos artigos e modelos. Algumas das atividades como reuniões no metaverso, projetos audiovisuais imersivos e tridimensionais são descritas nos capítulos e são inspiração para os tópicos deste estudo. Para este grupo foi idealizada uma identidade visual que ocupou todo o processo de pesquisa, iniciando em websites de registro, assinatura de eventos e até a publicação.

¹ Disponível em <https://visgraflab.impa.br/xr-ai>



Figura 1 - Grafismo criado para o grupo de estudos pensado como responsivo, isto é que pudesse ser flexível para aplicações em diferentes situações fonte: próprio autor

O grafismo que compõe a identidade visual (Figura 1) busca comunicar a interseção de ideias, a noção espacial e o sistema de malhas. Criado em 2022, durante o início dos encontros do grupo. Destas atividades, podemos observar ainda uma multiplicidade de formas de representação em experiências de realidade expandidas. Fazendo o cruzamento com plataformas, formatos e fluxos de desenvolvimento, é possível estabelecer uma base para observação da evolução de visualizações de mídia tridimensional.

O recorte desta pesquisa prioriza a análise e criação de imagens para experiências tridimensionais, seu tratamento, método de captura, modelagem, estrutura e até mesmo pertinência diante da representação desejada como instrumento de comunicação e divulgação. Aqui trabalhamos como finalidade a representação de objetos tridimensionais de interesse histórico-cultural. Na perspectiva do design, estes podem ser definidos como artefatos, edificações ou conjuntos materiais cuja forma, volumetria, materialidade, técnicas de fabricação e modos de uso concentram valores reconhecidos socialmente (históricos, estéticos, técnicos e simbólicos) e, por isso, demandam decisões projetuais de mediação para sua documentação, preservação, comunicação e fruição pública. Nesse enquadramento, tais objetos se aproximam do patrimônio cultural material (monumentos, conjuntos e sítios), mas também se articulam ao patrimônio imaterial quando seu valor depende de saberes e práticas (ofícios, rituais, “saber-fazer”) que sustentam sua produção e significado (UNESCO, 1972; UNESCO, 2003). Do ponto de vista do projeto, o “interesse” do objeto não está apenas em sua forma, mas na sua condição relacional: escala e ergonomia, marcas de uso, circulação, contexto e proveniência — aspectos que exigem lidar com tensões como acesso versus conservação, legibilidade versus precisão, e interpretação versus evidência, tornando o trabalho um

problema complexo de escolhas (Buchanan, 1992), escolhas tanto do material a ser representado como sob qual ótica e ângulos.

Durante atividades do grupo Space-XR, estudamos por exemplo a possibilidade de alguns usos de metaversos. Este nome, cunhado por Neal Stephenson no livro *Snowcrash* (2021) e passou a denominar espaços sociais virtuais que operam utilizando tecnologia tridimensional. O termo foi apropriado pela empresa Meta, porém podemos pensar mais conceitualmente nesta configuração propriamente. Apesar dos estudos serem conduzidos em boa parte utilizando HMD Metaquest 2 e 3, aplicativos como o Horizon, que pretendia ser um “Metaverso”, não foram utilizados porque suas especificações não atendiam a algumas premissas do grupo. Após um relatório técnico comparativo entre plataformas disponíveis (Velho et al., 2021), plataforma Spatial.io foi escolhida pelo grupo pela diversidade de possibilidades de acesso, riqueza de personalização e agilidade de produção, o que facilitava iterações. Características elencadas considerando a prototipagem como parte da ação em estudo de novas mídias.

Na continuidade de alguns experimentos realizados grupo Space-XR, observamos uma mudança de paradigma na representação tridimensional: a possibilidade de utilização de campos de radiância, isto é, possibilidades de representação tridimensional sem o uso do sistema de malhas e texturas. Em termos simples, campos significa espaço e radiância, cor. Adotando como observação alguns processos do grupo Space-XR, criamos alguns tópicos que servem de base para experimentos. Estes pretendem expandir as possibilidades de aplicação, representação e processos. Ao cruzar esses métodos em boa parte da pesquisa, temos uma multiplicidade de métodos e ferramentas a serem avaliados.

Essas novas formas de representação, são principalmente NeRFs (Mildenhall et al., 2020) e Gaussian Splatting (KERBL et al., 2023), associadas a uma abordagem por campos de radiância. Estas são representações que modelam a aparência da cena como função de posição e direção de observação, e a partir de elementos visuais distribuídos no espaço, viabilizam síntese de novas vistas. Nesta pesquisa, adota-se o Gaussian Splatting (3DGS) como forma preferencial de representação, em razão de seu potencial experimental, diversidade de aplicações em desenvolvimento e de sua viabilidade para

renderização eficiente. Mesmo que estes formatos evoluam, algo natural diante da velocidade do avanço tecnológico, o campo de radiância como conceito de representação de imagem tridimensional parece bastante promissor a longo prazo.

Os sites [radiancefields](#) e [supersplat](#) (“SuperSplat - The Home for 3D Gaussian Splatting”, [S.d.]) são fundamentais sobre o uso e desenvolvimento de 3DGS neste um momento dinâmico, quando coletivamente se aprende sobre as possibilidades do formato. Além do repertório técnico, é possível ampliar o repertório em aplicações em casos e utilizações em XR e diferentes mídias.

Em relação aos processos, usamos a teoria da caixa preta de Flusser (2011) para pensar sobre o processamento da imagem tridimensional e sobre inteligência artificial. Observando cada processo como uma caixa que pode ser aberta ou fechada, variando assim a possibilidade de reconfiguração e uso, comparamos com programas como Comfy Ui (“ComfyUI | Generate video, images, 3D, audio with AI”, [S.d.]).

Mesmo adotando a técnica de representação para campos de radiância, muitos processos utilizando malha geométrica e textura são utilizados, dado o repertório já desenvolvido. Dadas as diferenças, podemos adotar como pressuposto que uma forma de representação não substitui a outra, e as características de cada uma são discriminadas no decorrer da pesquisa.

Quase em contraposição em relação a adoção de uma representação tridimensional direta que é o 3DGS, um recurso que atravessa a pesquisa é a possibilidade de utilização de aprendizado por máquina, visão computacional e inteligência artificial, especialmente para observar e gerar imagens, preencher lacunas de informação e gerar tridimensionalidade a partir de imagens bidimensionais. A visão computacional pode vir a ampliar possibilidades de design, assim como a inteligência artificial generativa, pautada por referências e treinamento pode colaborar em referências. Nesse sentido iniciamos pelo bidimensional para observar treinamento, e realizamos testes e experimentos utilizando ferramentas tridimensionais, de vídeo e de visão computacional para produzir tridimensionalidade.

Observamos características das experiências criadas pelo grupo em análise empírica e com base em modelo teórico para estudo de interatividade apresentado no capítulo Experiências e Ambientes Interativos (Mitchell; Novaes, 2024), que sintetiza ferramentas de análise para ações de projetos para experiência. O modelo exposto traz uma tipologia que esclarece possíveis características de experiências, a começar pela interatividade sensório/corporal que um pressuposto da utilização do HMD.

A IA generativa de imagens e vídeos resulta da convergência entre avanços no processamento de linguagem natural, na modelagem generativa e na compressão de representações visuais. Nesse arranjo, arquiteturas do tipo *transformer* codificam prompts em representações semânticas capazes de orientar a síntese visual (VASWANI et al., 2017), enquanto os modelos de difusão passaram a oferecer um procedimento para gerar imagens a partir da remoção progressiva de ruído (HO; JAIN; ABBEEL, 2020). A adoção do espaço latente tornou esse processo mais eficiente, ao deslocar a geração do domínio direto dos pixels para representações comprimidas, preservando qualidade visual com menor custo computacional (ROMBACH et al., 2022).

No caso do vídeo, essa mesma base técnica é estendida para incorporar coerência temporal entre quadros, fazendo com que os sistemas generativos passem da produção de imagens estáticas para sequências audiovisuais consistentes, o que evidencia que a IA generativa contemporânea não depende de uma única técnica isolada, mas da articulação entre compreensão semântica, modelagem probabilística e representação latente (HO et al., 2022).

Testamos na prática a especificidade dos objetos e cenários em contraposição à imagem generativa, inscrições, texturas e poses. Do ponto de vista de representação tridimensional, são estabelecidas novas propriedades, possibilidades e características a partir da implementação do 3DGS como forma de representação. Na perspectiva de utilização de inteligência artificial, é impossível avaliar completamente as inovações em curso, no entanto, através da comparação com projetos já realizados, é possível contrastar e parametrizar alguns tópicos de abordagem. A contraposição entre documentação tridimensional, criação imagética e inteligência artificial, diante de tópicos consagrados por experimentos, possibilita reflexões acerca da imagem digital como um

todo e sobre a representação tridimensional em ambientes imersivos. Com isso, coloca-se a questão central:

Observando um repertório de experiências de conteúdo científico-cultural e seus processos, como utilizar representações por campo de radiância tridimensionais e processos de aprendizado de máquina para a construção de imagens para experiências imersivas?

Nessa perspectiva, este é um estudo que pretende contribuir com o repertório de representações visuais para a representação tridimensional em realidades expandidas, registrando casos e testando possibilidades, para construções de experiências e observar características de uso de IA em meio tridimensional. Além abaixo seguem objetivos secundários através dos quais procuramos elucidar a questão principal.

- Observar as características técnicas de geração de imagens em Gaussian Splatting e elencar diferenças da representação entre modelos tridimensionais em malha e campos de radiância no âmbito de experiências imersivas.
- Investigar a inteligência artificial, aprendizado de máquina e visão computacional como ferramenta para geração de imagens tridimensionais
- Avaliar a criação de ambientes imersivos, considerando a digitalização de ambientes reais e a criação de ambientes fictícios.
- Observar as possibilidades de representação humana em campos de radiância e possibilidades de captura de movimento.

Esta pesquisa tem origem em uma trajetória situada entre **prática experimental** e **reflexão crítica** sobre a construção de experiências virtuais tridimensionais imersivas. Parte-se, em especial, da participação em projetos do grupo de estudos **Space-XR**, bem como de estudos de caso derivados dessas realizações, articulados a atividades docentes na **PUC-Rio** nas disciplinas de Realidades Digitais e imagem, corpo e moda.

Esta é uma pesquisa qualitativa, que utiliza-se do método de pesquisa através do design, tal qual Frayling (1993) aponta, em prática reflexiva (Schön, 1983) utilizando

tópicos especificados por Godin e Zaedi (2014). Esta prática reflexiva fomenta a revisão bibliográfica sobre documentação de processos correlatos, assim como documentação teórica em várias mídias e periódicos sobre tecnologia. É uma pesquisa que busca extrair teoria e metodologias através da prática do design.

Do ponto de vista da experiência, os projetos analisados através de prática reflexiva frequentemente propõem o deslocamento de objetos e/ou ambientes para um regime de fruição digital mediado por dispositivos imersivos. A partir do acompanhamento de processos de XR baseados em malha e em vídeo/360°, foram iniciados experimentos com geração de imagens por modelos generativos, representações por campos de radiância, com ênfase em Gaussian Splatting, e técnicas de visão computacional para captura de movimento sem sensores e estimativas de profundidade, buscando ampliar a articulação entre imagem, corpo e espacialidade interativa.

A inteligência artificial, aprendizado de máquina e visão computacional nesse caso são alguns processos apresentados nos capítulos. No momento, além de modelos de geração de imagem e texto, existem modelos de geração tridimensional e os *worlds models*. Fazem parte de fluxos propostos e aparecem como parte dos recursos existentes.

O design de experiências em Realidade Virtual (VR) e Realidade Expandida (XR), entendida como guarda-chuva que abrange VR, AR e MR ao longo do continuum realidade–virtualidade, depende de um equilíbrio delicado entre percepção e viabilidade. Por isso, o esforço em tornar factível a percepção de realidade, principalmente nestes casos em que os objetos apresentados virtualmente possuem importância histórico-culturais.

Essa escolha implica uma mudança relevante tanto no método quanto no resultado na representação, já que a construção deixa de depender primariamente de modelagem e passa a se organizar pela otimização e composição de primitivas volumétricas que reconstituem a aparência em função do ponto de vista (KERBL et al., 2023). Nesse contexto, métodos recentes de **campos de radiância** reconfiguram a forma de repre-

sentar cenas: em vez de descrever prioritariamente objetos por geometria explícita, descrevem **aparência e emissão de luz** de modo consistente com a variação de ponto de vista. A utilização de imagens tridimensionais com aparência fotográfica pode facilitar a percepção de imediação (Bolter; Grusin, 2002), isto é, esquecer a mediação, através do sentimento de presença.

Representações tridimensionais podem interessar ao design de experiências expandidas, uma vez que afeta a experiência e por isso seu estudo contribui para recursos visuais narrativos, processos e diferentes perspectivas. O estudo aborda representação a partir de novos paradigmas construtivos e métodos que mesclam.

No caso de experimentos de deslocamento para o virtual, de objetos e ambientes, isto é essencial no sentido de tornar verossímil a experiência. Por outro lado, aspectos específicos às vezes aparecem, como a nitidez de texturas ou representação de sombras. Não se trata de diminuir a importância ou a pertinência do uso da estrutura de malhas e texturas, mas de caracterizar diferenças entre aplicações e processos. Alguns itens como pelos, reflexos, plantas, transparência e até emissão de luz podem ser capturados por este método. Em VR/XR, a **presença/telepresença**, a sensação de “estar lá” se sustentam pela coerência sensorio-motora e pela vividez do ambiente.

Experiências XR frequentemente dependem de microtexturas, irregularidades e variações sutis de iluminação para produzir convicção perceptiva. Ao privilegiar uma representação de aparência capturada que sustenta variação de ponto de vista, o Gaussian splatting tende a favorecer uma presença “fotográfica” em cenas reais que atendem a representação de patrimônio, museus, ambientes urbanos e representação humana.

Por outro lado, como uma nova técnica de representação, é interessante observar suas características gráficas em sua função projetada e experimentais, desviando de sua função primária, visto que possui elementos construtivos e expressivos.

No *pipeline* tradicional, alcançar realismo exige a captura de origem semelhante, mas depois disso é gerada a malha, UV, baking, materiais PBR. Já em *pipelines* baseados em *radiance fields*, o centro da autoria desloca-se para qualidade de captura, lim-

peza/otimização e direção de experiência (ritmo, pontos de interesse, narrativa espacial). O 3DGS é particularmente relevante por mirar renderização eficiente, mesmo que para realizar cenas amplas tenha um custo de processamento gráfico elevado.

No capítulo 2, Luz, cor e espaço apresentamos ao 3DGS, descrevemos fluxos de captura, treinamento e edição da técnica de representação, com alguns exemplos. Realizamos ainda uma comparação ao sistema de malhas e texturas e o 3DGS. Refletimos sobre aspectos visuais comparando com a fotografia.

No capítulo 3, trazemos experiências do grupo Space XR entre digitalização, tratamento e apresentação de objetos de interesse histórico-cultural e exposição no Metaverso. Observamos a captura de ambientes para imersão, relacionando estas experiências a experimentos em 3DGS. Registramos o processo de distribuição de imagens e modelos tridimensionais, em AR, VR e diferentes suportes tecnológicos, avaliando a importância de narrativas e preparo de experiência. Exploramos casos relativos a artefatos como Sha Amun em Su, a ambiência do projeto Neferhotep 360, a implementação de modelos e metaversos em ambiente web, e a transposição do projeto V-horus.

No capítulo 4 apresentamos produções baseadas em ilustrações de capoeira, dança e objetos de interesse histórico-cultural utilizando inteligência artificial generativa e visão computacional. Trazemos experiências através de 2D, 3D e movimento, experimentando diferentes métodos, como modelos de geração tridimensional e criação de vídeos para modelagem.

No capítulo 5 apresentamos especificamente a representação humana e movimento. Partimos de questões referentes a vale da estranheza (Mori; MacDorman; Kageki, 2012) e observação de casos. Registramos processos de captura de movimento (mocap), avatares em realidade virtual e suas características para pontuar a utilização de 3DGS e 4DGS, apontando caminhos alternativos para a criação de 4DGS, utilizando inteligência artificial.

Por fim avaliamos a relação entre 3DGS e modelos tridimensionais em malha, a produção de imagens através de inteligência artificial e questões do design em relação aos métodos diante das novas tecnologias.

2.

Luz, cor e espaço

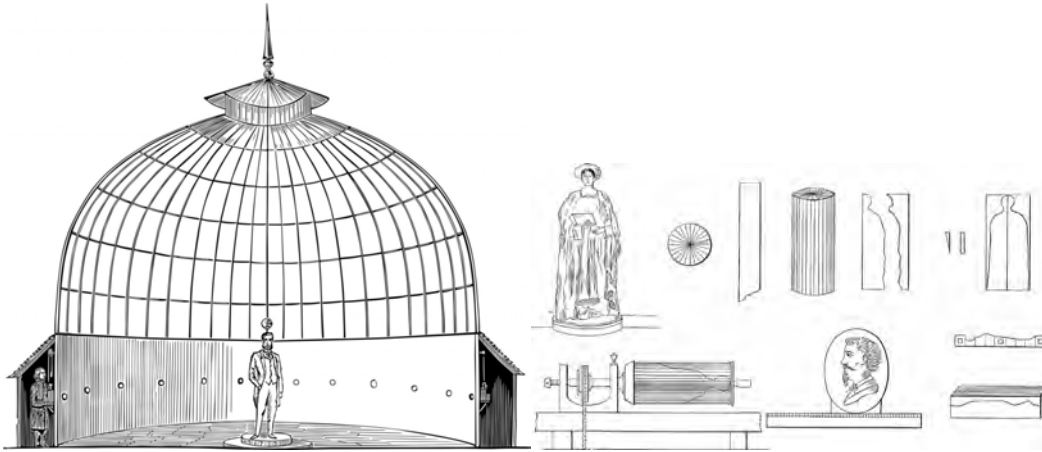


Figura 2 - Fotoescultura de François Willème. fonte:(Laric, [S.d.]

Observar em três dimensões a partir de fotografias remonta a François Willème (1830–1905), escultor francês que, a partir de 1859, desenvolveu a *photosculpture* (fotoescultura) um método pré-computacional de “captura 3D” feito com múltiplas fotografias ao redor do corpo/rosto para produzir retratos escultóricos tridimensionais (SOBIESZEK, 1980) como método e assim, operando, na prática, uma reconstrução volumétrica a partir de múltiplas vistas, conforme figura 2, tal qual um parente conceitual de técnicas modernas como fotogrametria, só que mecânico e manual (Park, 2025).

Também é possível buscar as estereoscopias, fotografias feitas com duas lentes que transmitem a sensação de tridimensionalidade, como busca da sensação do tridimensional. Vertente da fotografia tradicional criada por Charles Wheatstone em 1860 e popularizada por David Brewster, provoca percepção do tridimensional ao expor imagens levemente diferentes, em técnica semelhante à dos filmes 3D atuais. Podemos

então perceber que há dois caminhos que se cruzam, um em mensurar a tridimensionalidade para reprodução e outro em percebê-la oticamente. Embora muitas vezes funcionem juntas no ambiente digital, a discriminação dessas intenções pode colaborar com a distinção de finalidade, a partir de sua estrutura. Enquanto na primeira é importante a forma total e precisa, na segunda existe o ponto de vista como eixo da experiência e a aparência fotográfica e realidade como forma final.



Figura 3 -Alfred Silvester, Pequena Nell. Em – 'A Loja de Antiguidades' de Charles Dickens, década de 1870-1880. Duas impressões em albúmen em formato estereoscópico. Fonte: Coleção de Artes Gráficas GAX



Figura 4 - Lenticular or 'Brewster' stereoscope, onde é possível notar a semelhança com headsets XR
Fonte: ("180 years of 3D | Royal Society", 2018)

Podemos notar que, se por um lado não há nenhuma volumetria no sistema lenticular, o sentido ótico, juntamente com a exposição dos dois negativos, é o que traz um índice de realidade, através da disparidade binocular. O índice de realidade gravado em filme poderia produzir uma sensação de profundidade. Mas devemos considerar que o esquema de Willème tinha a informação do volume, não intenção ótica, essa seria produzida pela própria escultura, ao final do processo. Considerando a utilização da silhueta, as fotoesculturas são um parente conceitual de técnicas modernas como fotogrametria, só que mecânico e manual (Park, 2025). Já a fotografia estereográfica corresponde a vídeos estéreos e 360, ao próprio *headset* de XR. O pensamento de campo de radiância parece o encontro das duas técnicas. A técnica de Willème influencia completamente o processo de fotogrametria, mas trabalha com superfície. A fotografia estereográfica, por sua vez, é capaz de entregar as nuances fotográficas, matriz na qual a complexidade e sutileza são dependentes da granulação da qual a imagem é formada.

O conceito de campo de radiância — isto é, a ideia de representar a luz emitida por uma cena em função da posição e da direção — remonta aos estudos de Levoy e Hanrahan (Levoy; Hanrahan, 1996) e de Gortler et al. (1996), que introduziram o conceito de *light field* como um conjunto denso de amostras de raios de luz. Nessas abordagens, a geometria tradicional baseada em malhas era secundarizada em favor de uma codificação da cena como campo de aparências. A genealogia conceitual de remonta aos *light fields* (Levoy; Hanrahan, 1996) e ao Lumigraph (Gortler et al., 1996), que tratam a cena como amostragem do fluxo de luz e permitem reamostragem para gerar novos pontos de vista, ainda que com limitações práticas e exigências de captura densa. Na época do lançamento, o poder computacional era limitado para esse método. O tema foi desenvolvido por Debevec e Malik (1997) que chegaram a influenciar a cinematografia com o bullet-time, efeito criado para o filme Matrix (1999), fabricado de maneira semelhante a captura tridimensional, com múltiplas câmeras ao redor do personagem.

Algum tempo depois, é elaborado o EWA Volume Splatting (Zwicker et al., 2001), que propõe *volume rendering* por *splatting* baseado em kernels gaussianos elípticos e em um filtro de reamostragem (combinando reconstrução e um passa-baixas) para reduzir artefatos de aliasing sem introduzir borramento excessivo, em analogia ao filtro EWA (*elliptical weighted average*) discutido na literatura de filtragem (ZWICKER et al., 2001).

Em 2018 a Google, através de pesquisas de Paul Debevec, chegou a lançar uma aplicação para a plataforma Steam para observação em realidade virtual, que tinha capturas em light Fields (“Welcome to Light Fields no Steam”, [S.d.]). Na página Experimenting with Light Fields (“Experimenting with Light Fields”, 2018) aparece um rig em arco rotativo de 16 câmeras para captura.

Essa linhagem foi retomada e reformulada com o advento dos *Neural Radiance Fields* (NeRFs), que modelam a cena como uma função contínua parametrizada por redes neurais (Mildenhall et al., 2020). No entanto, apesar de sua qualidade visual, os NeRFs enfrentam limitações práticas, especialmente quanto ao desempenho em tempo real e à dificuldade de generalização para dispositivos com poder computacional limitado.

A partir desses elementos, surge o Gaussian Splatting, como proposto por Kerbl et al. (2023). O 3D Gaussian Splatting surge como um “ponto de equilíbrio” ao representar a cena por um conjunto de Gaussianas 3D anisotrópicas, isto é, muda de aparência conforme o ponto de vista, otimizadas e renderizadas por um algoritmo rápido de splatting/visibilidade, buscando alta qualidade com renderização em tempo real em cenários relevantes (Kerbl et al., 2023). O 3DGS, teve ampla aceitação por conta do design, simplicidade de implementação e pertinência.

Um ponto importante é a compatibilidade com processos anteriores, na medida que possibilita o uso do conjunto de imagem realizado para fotogrametria. Deste modo, é possível herdar dados e processos de captura.

‘v2.1. Captura e treinamento

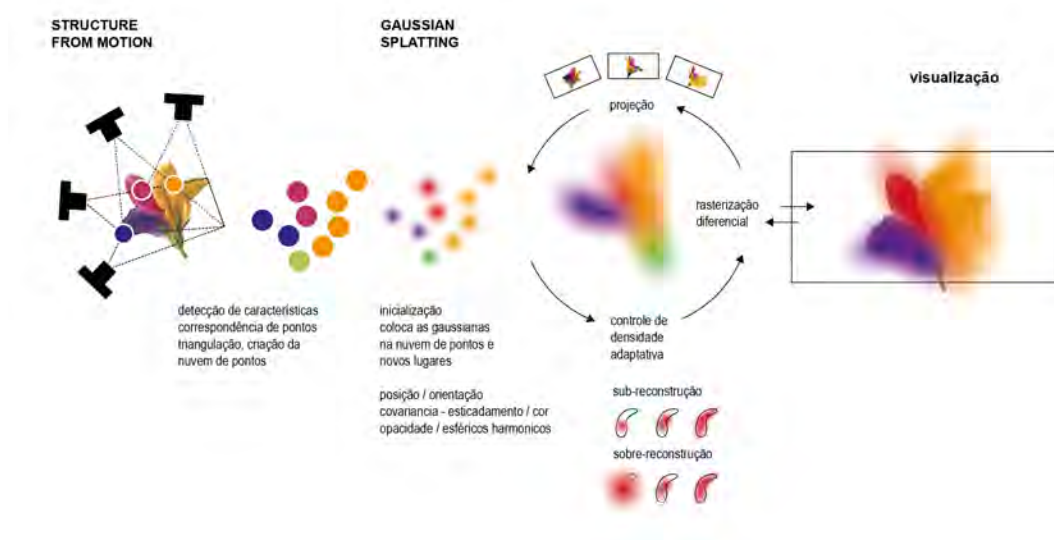


Figura 5 - Gráfico do fluxo de treinamento básico de 3DGS

O processo inicia-se, em geral, pela **captura fotográfica ou por vídeo**, convertendo-se o vídeo em quadros amostrados, desde que haja sobreposição entre imagens, variação suficiente de ponto de vista (paralaxe) e condições de captura que preservem nitidez e estabilidade fotométrica. Em geral 200 a 500 imagens para um objeto são

suficientes. A sobreposição entre as imagens é muito importante para o treinamento, assim como nitidez e continuidade. É sempre interessante lembrar que a relação entre visualização e captura é interdependente, isto é, se a vista de cima será observada, ela deve estar planejada na captura. O *shutter* alto, em torno de 250 e movimentos lentos são recomendáveis para evitar desfoque de movimento. Da mesma forma, é recomendável o diafragma fechado, algo em torno de f8-f11 é importante pois o desfoque de profundidade faz com que a imagem perca definição. Por isso, a utilização do smartphone para captura pode não ser precisa quanto o uso de uma câmera onde as configurações são ajustadas especificamente para captura.

As diferenças entre a captura através de fotos e vídeos são as das próprias imagens capturadas. As fotos possuem mais nitidez e resolução enquanto o vídeo obtém muitas imagens rapidamente. Levando-se em conta o tempo de captura, para objetos dinâmicos como, por exemplo, pessoas, que involuntariamente se movem mesmo quando procuram manter-se estáticas, o vídeo se torna uma boa ferramenta; mas para ambientes e objetos, preferencialmente podem ser registrados através de fotos.

Em seguida, aplica-se um estágio de Structure-from-Motion (Ullman, 1979), no qual correspondências entre pontos de imagem são detectadas e reconciliadas entre diferentes vistas para recuperar parâmetros intrínsecos ou extrínsecos de câmera (posições e orientações, como na figura 5), além de uma nuvem de pontos esparsa que serve como estrutura inicial da cena (Schonberger; Frahm, 2016). Essa etapa é decisiva porque fornece o “espaço de referência” no qual as gaussianas serão colocadas e, sobretudo, porque define o conjunto de projeções que a otimização posterior deve satisfazer. Esse processo é comum ao processo de fotogrametria tradicional, e operado por sistemas como colmap (Schonberger; Frahm, 2016).



Figura 6 - Structure from motion- Mecanismo que observa pontos de interseção na imagem e posiciona em um espaço tridimensional, criando uma nuvem de pontos e o mapeamento dos pontos de vista na extração de colmap.

Com as poses estimadas, a representação por gaussianas é **inicializada** a partir da geometria esparsa, atribuindo-se a cada elemento parâmetros de posição e forma, bem como termos de aparência que podem incluir dependência direcional da visão. A etapa central consiste na **renderização diferenciável**: cada gaussiana 3D é projetada para o plano de imagem como uma distribuição 2D elíptica e suas contribuições são compostas por mistura alfa para sintetizar uma imagem predita; em seguida, minimiza-se uma função de perda que quantifica a discrepância entre a imagem renderizada e a imagem capturada correspondente, atualizando-se os parâmetros das gaussianas via gradientes (KERBL et al., 2023).

Durante o treinamento, o algoritmo executa as seguintes etapas ao longo de milhares de iterações: adiciona *splats* em áreas que precisam de mais detalhes, remove manchas que não contribuem significativamente e ajusta os parâmetros de dispersão para melhor corresponder às imagens de entrada (figura 6).



Figura 7 - breve seqüência do treinamento em 3dgs. Disponível em

A cada 100 gerações, a imagem é redesenhada. A cena final, portanto, resulta de um compromisso entre fidelidade às imagens observadas, estabilidade geométrica derivada das poses e eficiência computacional, produzindo uma cena navegável sob novos pontos de vista por projeção e composição das gaussianas aprendidas (KERBL et al., 2023).

Existem diferentes maneiras de se obter 3DGS, algumas mais controláveis e complexas de operar e outras mais simples e diretas, sem que haja muita interação no processo. O código base para treinamento e visualização provido na própria publicação original é suficiente, porém requer uma instalação através de código.

O Luma interactive scenes (“Luma AI - Interactive Scenes”, [S.d.]) foi pioneiro ao implementar uma interface móvel para treino de imagens para 3d, com saída para 3DGS e outros formatos como glb. Foi seguido pelo Polycam, aplicativo de fotogrametria em nuvem. Ambos oferecem maneiras extremamente simples de captura e treino e operam o processo em nuvem e quase invisível. O Kiri Engine (Engine, [S.d.]) opera de maneira semelhante, com algumas opções a mais, como eliminação do fundo. Estes aplicativos funcionam tanto através do celular, quanto a partir do envio de imagens / vídeo. Para mais controle e processos mais complexos são indicados programas de treinamento local que em geral necessitam de placa de vídeo para processamento, o programa proprietário Jawset Postshot (“Jawset Postshot”, [S.d.]) que oferece um fluxo plenamente personalizável a partir do conjunto de imagens ou vídeo, até o 3DGS final, além do Nerfstudio. O Brush e Lichtfeld Studio são alternativas de código aberto. Estas necessitam da etapa de extração de colmap, que pode ser realizada no próprio colmap

ou no Epic Reality Scan. O Lichtfield tem incorporado técnicas adicionais no treinamento de splats como Bilateral Guided Radiance Field Processing (“Bilateral Guided Radiance Field Processing”, [S.d.]; Wang et al., 2025) e abrirá para plugins externos. Existem muitos outros programas e sistemas, como Gsplat e pode ser executado no Nerfstudio, mas que requerem algum conhecimento técnico para sua instalação.

Os principais parâmetros são número de splats, em geral entre 100.000 e 3.000.000 e o número de iterações que por padrão são 30.000 no Jawset Postshot e no Lichtfield Studio. Um maior número de iterações não garante melhor qualidade, assim existe um balanço entre complexidade da cena, número de splats, e número de iterações, que pode ser de alto detalhe para cenas externas, ou bem menor para objetos, por exemplo. O ideal é avaliar observando o treinamento, visível na maioria dos programas.

Podemos citar ainda algumas técnicas adicionais. Marcadores April tags podem ser utilizados para aumentar a precisão no registro das imagens. April tags são formas fiduciais, símbolos gráficos semelhantes a QR codes que foram criadas para auxiliar em visão computacional em robótica. São marcas de referência de alta precisão que indicam posição e escala e são processados durante o processo de extração em *surface from motion*.

A utilização de máscaras é um recurso útil tanto para remoção de objetos e pessoas e fundo, quanto para especificar a área a ser treinada. Para detecção de itens é possível utilizar visão computacional, em geral YOLO ou DINO. A especificação da área treinada pode ser feita no Reality Scan através da geração de um modelo em malhas com o objetivo de estabelecer superfície ou através da detecção de objetos através de IA. Com esse recurso, torna o treinamento objetivo e impede a criação de splats no entorno do objeto, diminuindo a necessidade de retoque.

2.2. Edição

Os arquivos gerados muitas vezes são um pouco densos, em um primeiro momento aparecem muitos splats em torno da área ou objeto de interesse. Porém, após

entrar ou organizar a cena, navegando nela, é possível chegar ao núcleo de interesse da captura e dessa forma visualizar melhor a imagem. São poucos os programas e plataformas para edição de 3DGS. Cada programa de treinamento conta com módulos de edição, porém o mais popularizado é o Supersplat, um aplicativo online. Este é um editor *open source* implementado na plataforma playcanvas, que agrega ferramentas de limpeza e transformação.

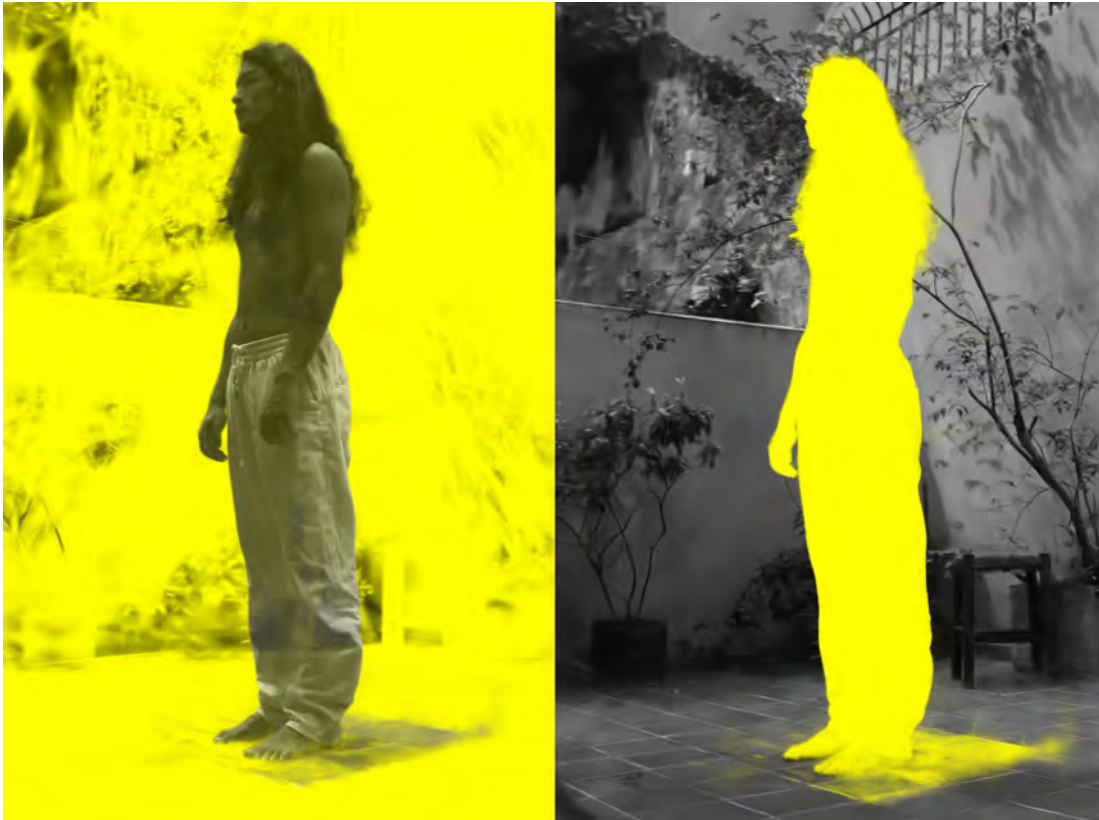


Figura 8 - Seleção realizada no Supersplat em 3DGS, separando a pessoa representada, João Vitor Silva, do ambiente fonte:

A seleção por caixa, desenho ou forma geométrica são de simples aprendizado. Talvez a maior qualidade na edição no supersplat seja a seleção através das elipsóides visíveis e selecionáveis, facilitando assim a “limpeza” das cenas, como na figura 7. A presença de ferramentas como brilho e contraste, temperatura de cor e tom nos remete aos recursos de ajustes de imagem presentes no photoshop. Além disso, a possibilidade de incluir várias “cenas” cria a possibilidade de composições. A tridimensionalidade

colabora na compreensão do material capturado principalmente quando é intrínseco, pois é possível selecionar a área de interesse, inverter a seleção e apagar as áreas residuais em torno. A edição de ambientes é mais complexa pois a sobreposição de splats faz parte da construção das cenas e a alta densidade e a distribuição tornam um pouco mais difícil de separar ou limpar.

Existem duas dificuldades centrais: ao editar e apagar uma área é comum sobram splats, além disso, em muitos casos, existem splats flutuantes nas cenas. Para o primeiro caso, existe uma ferramenta em desenvolvimento pela Nvidia chamada ArtisanGS (Tsang et al., 2026), que propõe uma seleção por detecção de objeto adaptando a superfície e apara os splats de objetos automaticamente.

Existem muitos recursos distribuídos em diferentes programas, selecionar por tamanho de elipsóide, por exemplo, postshot, alterar a iluminação de um objeto ou cena, que é algo simples em 3D por exemplo, ainda é um recurso restrito ao Volinga, um *plugin* do Unreal Engine, mas algo desejável que pode vir a acrescentar recursos visuais e de linguagem.

2.3. Uma breve comparação e características

Desde a década de 70, um amplo repertório de ferramentas para representação tridimensional se desenvolve baseada em malhas e texturas. A malha é o que conhecemos como estrutura onde aplicadas texturas e exposta a luz, revela-se uma simulação de forma. A partir desta estrutura são modelados cenários, objetos e colocados em movimento através de uma série de recursos como física, esqueletos e mecânica.

À exemplo do sistema de captura de movimento descritos no capítulo 4, toda uma lógica entre malha, texturas e estruturas foi construída em longa trajetória. A representação por campos de radiância foi concebida também há bastante tempo, mas não foi desenvolvida porque envolvia um processamento complexo, hoje facilitado pelo desenvolvimento das placas gráficas.

Não cabe elencar uma forma melhor, mas observar as características. Muitos casos de desenvolvimento de projetos XR hoje seguirão utilizando modelos 3D em malhas simplesmente pela integração em fluxos de processos. Cabe aqui elencar características e a comparação pode facilitar a compreensão de aplicação em cada caso e as diferenças.

O ecossistema de ambientes virtuais, metaversos, modelos e programas têm fluxos de trabalho estabelecidos e altamente desenvolvidos. Um bom exemplo é complexidade de um *Metahuman* sistema da Epic desenvolvido para representação paramétrica humana, com o objetivo de representação realista além do vale da estranheza (Mori; MacDorman; Kageki, 2012), ou mesmo, a percepção de alternativas de caminhos existentes em um cenário virtual através de inteligência artificial, muito também influenciados pela indústria audiovisual, traz possibilidades de animação em render, efeitos especiais. Em tempo real, cenários de cinema são produzidos em *virtual production*, onde o cenário digital em led é filmado em estúdio como uma *cave*, considerando registro com a posição da câmera, para essa última.

Enquanto a malha (coberta por uma textura) representa a cena/objeto como uma **superfície**: vértices, arestas e faces (triângulos) + uma imagem de **textura 2D** (UV) que “embrulha” a imagem na geometria como pode ser visto na figura 8.

Surface from Motion (Ullman, 1979) → MVS denso → reconstrução de superfície como, por exemplo, Poisson (Kazhdan; Hoppe, 2013) → texturização/UV (Cattull)



Figura 9 - Estrutura de modelo tridimensional / em baixo número de polígonos para deixar a estrutura visível Fonte: Próprio autor

O modelo em malha é mensurável em termos de topologia, e pode distinguir partes e superfície precisamente em sua modelagem. Por esse motivo, é indicado para impressão e racionalização da forma tridimensional. As representações determinísticas com vértices e polígonos definidos (como malhas triangulares) facilitam a obtenção de formas reais e virtuais. Essa forma do modelo em malha pretende ser o mais definida possível, sendo ideal que o objeto ao final do processo seja fechado, sem buracos, com a superfície mapeada. Para objetos modelados geometricamente com malhas, temos simulação de gravidade, densidade e todo tipo de recursos, a exemplo de estrutura como *rigs*, criando articulações para personagens ou mecanismos, para objetos. Em especial para representação, temos texturas normais que simulam microtexturas, preservando a geometria leve.

A nuvem de pontos utilizada de maneira universal é criada a partir do encontro de características das imagens identificáveis enquanto se movem. A partir da mudança durante o movimento a forma é sugerida. Desse modo, resulta em uma sugestão da forma através da distribuição de pontos no espaço tridimensional. O ponto é, como diz

Kandinsky (2005), o momento inicial de força interior e exterior. Ao ver a representação dos pontos, apreciamos sua força exterior. Os sistemas, ao utilizar a nuvem de pontos, utilizam as propriedades interiores dos pontos, demarcando o espaço.

A captura 3D “por malhas” e por Gaussian Splatting partem, muitas vezes, do mesmo insumo, conjunto de fotos/vídeo com poses de câmera estimadas via *structure from motion* (como descrito acima), mas terminam em representações bem diferentes: a malha privilegia geometria explícita; o Gaussian Splatting privilegia aparência e cor (radiância), em geral, renderizável em tempo real.

Esta técnica permite que a geometria permaneça ambígua ou fluida, o que é útil para representar fenômenos visuais complexos como cabelo, folhagem, transparências e reflexos, regiões difíceis de modelar por malhas durante a captura (Kerbl et al., 2023).

O Gaussian Splatting (3DGS) representa a cena como um conjunto de primitivas volumétricas (Gaussianas 3D anisotrópicas) com parâmetros de cor/opacidade e um renderizador por “splatting” (rasterização acumulativa) com visibilidade.

Em geral: SfM (pontos esparsos + poses) otimização das Gaussianas para reproduzir as imagens → espaço com gaussianas.



Figura 10 - Estrutura de modelo em 3DGS / em baixo número de splats para deixar a estrutura visível
Fonte: Próprio autor

Do ponto de vista construtivo, o Gaussian Splatting (3DGS) se baseia em uma nuvem capaz de capturar nuances da forma e da aparência da cena sem depender de malhas ou texturas projetadas. Através da sobreposição das elipses, em geral achatadas em várias direções, chamadas de splats, tal qual uma retícula, forma volumes de cor. Internamente, as elipses possuem degradês gaussianos semitransparentes, que remetem à curva gaussiana Carl Friedrich Gauss).

Outro ponto é que a renderização por Gaussian Splatting é feita por composição aditiva em ordem de profundidade, sem a necessidade de traçado de raios explícito. Isso a torna compatível com renderização em tempo real, inclusive em dispositivos móveis ou aplicações web.

O Gaussian splatting não substitui completamente a malha tradicional, sobretudo quando o projeto exige interação física, colisões precisas, objetos manipuláveis e edição semântica fina. O ecossistema de programas para edição deste desenvolvimento é mais evoluído, por ter sido o caminho adotado pela representação tridimensional por décadas. Por isso, uma estratégia recorrente é o hibridismo: splats para cenário/ambiente (presença e atmosfera) e malhas para elementos interativos (interface de usuário em 3D, ferramentas), alinhando fidelidade perceptiva e controle de interação, assim como a utilização do 3DGS sintético, onde são renderizadas centenas de imagens para treinar as gaussianas a partir de modelos em malha. Nesta perspectiva, contrastar os dois formatos podem enriquecer a propriedade projetual de escolha.

Uma diferença a ser apontada é a relação com as vistas no gaussian splats. Enquanto em malha é gerado um modelo uniforme que se comporta de maneira determinada pela configuração dos materiais, a interdependência do ângulo de visão cria uma dinâmica onde a cada ângulo a imagem pode mudar, tanto em relação a reflexos quanto a outros movimentos. Então, o primeiro forma um modelo integrado, um objeto digital, sendo suas características alteradas pela luz do ambiente virtual ou na edição das propriedades do material. No segundo caso, no 3DGS, ele guarda as propriedades de visão, reflexão e cor das fotografias. Assim, revela reflexões ou luzes que incidem sobre o objeto no momento da foto de maneira quase literal, e isto pode variar conforme o

ponto de vista, através da dinâmica proposta pelas harmônicas esféricas, imitando a imagem.

Por isso, é interessante notar que uma vez que não há uma concretude no 3DGS, talvez a classificação como modelo tridimensional não seja adequada. Talvez a imagem tridimensional ou arquivo espacial, ou ainda imagem espacial, denomine melhor campos de radiância, uma vez que, de um modo geral, são cores distribuídas no espaço.



Figura 11 – Uma forma dependente do ponto de vista Fonte: <https://superspl.at/view?id=29f9b77d>

2.3. Aspectos visuais

A realidade virtual e aumentada pode ser relacionada aos pré-cinemas, não só por ser um veículo audiovisual como pelo estágio de desenvolvimento. Isto é, ainda que existam propostas de equipamento e programas, narrativas e alternativas de representação ainda não sedimentados. O 3DGS é um bom exemplo disso. Algumas instalações de pré-cinemas dialogam diretamente com RV e RA, pela impressão da tridimensionalidade, assim como a exploração de possibilidades. Um exemplo desse encontro é o trabalho de Oskar Messter, que criou um sistema de projeção chamado alabastra (Loew, 2013), uma projeção que simula tridimensionalidade, com sincronia de sons, variante

do Pepper's Ghost, mais conhecido sistema de projeção e ilusionismo. Este inventor e diretor produziu filmes ao redor de modelos vivos estáticos chamadas Akt-Skulpturen, que puderam gerar imagens tridimensionais em 3DGS, 123 anos depois.



Figura 12 - Imagem em 3DGS baseada em filme de 1906 de Oskar Messter que usa modelos-vivos para criar cenas de Adão e Eva (existe uma maçã na mão da modelo, de difícil visualização). Fonte

Para a comunicação visual é possível uma via comparativa para aproximar a técnica de Gaussian Splatting de pintura, ponto fotográfico e retículas de policromia. Nesse sentido, consiste em reconhecê-los como respostas históricas distintas a um mesmo problema: como produzir continuidade perceptiva (tons, volumes e atmosfera) por meio de unidades discretas (toques, grãos/pixels, pontos de tinta ou nesse caso “splats”). No campo da pintura, essa lógica é explicitada por práticas que enfatizam a decomposição cromática e a integração perceptiva do observador, como no neoimpressionismo, cujo programa técnico reivindica uma pintura sistematizada por princípios de óptica e percepção (Amory, 2004).

Visualmente, a técnica pode gerar imagens foto-realistas, com transições suaves entre volumes, brilho e cor. A ausência de malhas rígidas e a natureza probabilística das Gaussianas proporcionam uma sensação de organicidade, próxima da percepção visual humana. Isso faz possível a representação luz e sombra, água, reflexo, cabelos, translucidez.



Figura 13 - Os elipsoides formam massas de cor através da popularização da nuvem de pontos

A relação entre fotografia e o 3DGS pode se iniciar na percepção de que esse sistema segue a captura de luz e cor, o que nos remete ao nome dado por Hércules Florence em 1833 (Campinas, SP), no contexto de experimentos fotoquímicos acerca da fixação da imagem produzida pela luz no papel.



Figura 14 – Imagem do autor - Fotorrealismo é facilitado através da técnica 3DGS, disponível em: <https://superspl.at/scene/c7f14114>

Assim como fica mais evidente ao observar o trabalho de fotógrafos utilizando a técnica, dupla exposição assim como longa exposição, exercícios expressivos que caracterizam processos a semelhança a exemplo de Edmund Fraser (“Edmund Fraser Studio”, 2025) com longa exposição como na figura 14 ou Eric Pare (“Eric Paré - Montréal - bullet-time / light-painting / gaussian splat photographer”, 2025) com lightpainting reforçam essa analogia, assim como as imagens macro de Dany Bittel (“Dany Bittel, Macro Splats”, [S.d.]) na figura 15. Por outro lado, podemos apontar a representação exata, nítida, utilizando condições específicas como a busca de Simon Bethke (“simon-bethke”, [S.d.]) pela precisão. Destes trabalhos e de outros podemos também diferenciar da fotografia pela espacialidade e ausência de moldura. Literalmente exploram outra dimensão, mas com linguagem bastante influenciada por elementos fotográficos. A possibilidade de colagens tridimensionais e de exploração visual é capaz de ir do surrealismo a tratamentos de cor.



Figura 15 - Edmund Fraser Gaussian Splat

Do ponto de vista perceptivo, a suavidade visual resultante da sobreposição de gaussianas é análoga a técnicas clássicas de pintura e desenho, em que o traço ou o

borrão não busca contornos precisos, mas sim uma síntese da impressão visual por acumulação (de pinceladas por exemplo). Essa qualidade aproxima o Gaussian Splatting mais próximo da percepção e menos da modelagem exata.



Figura 16 - Dani Bittel 3DGS

Em uma perspectiva desviante da representação perfeita, assim como podemos relacionar a coleta das imagens ao interesse dinâmico sobre a imagem como o faz David Hockney em suas colagens, podemos avaliar na perspectiva de forma, abstraindo da função objetiva figurativa, relações com pintura, através dos splats e da relação figura e fundo, desviando da representação perfeita (Li et al., 2025). A forma como o 3DGS ilustra o inacabado representa bem um ditado onde o defeito é um efeito, pois

embora as elipses longas das áreas com poucas informações possam ser um pouco agressivas, possuem propriedades expressivas.

Em síntese, representa um ponto de inflexão na forma como compreendemos a visualização tridimensional, traz propriedades de representação, o espaço e possibilidades de abstração diferentes. Talvez a própria reflexão sobre o objeto para documentação, anteriormente ligado a estátuas, esculturas e objetos capturáveis, através desta nova abordagem possa se dirigir a vegetais, pessoas, animais, formas orgânicas, translúcidas e ambientes com sua ambiência preservada.



Figura 17 - Imagem do autor apresenta a imagem incompleta, com splats visíveis como parte da linguagem ².

² Disponível em

3.

Entre o intrínseco e o extrínseco

A câmera intrínseca, apontada para o centro de um eixo e a câmera extrínseca apontada para o algo oposto ao eixo remontam ao Apple Quicktime VR (SIGGRAPH (Conference 22nd, p29: 1995 : Los Angeles et al., 1995), um dos primeiros programas imersivos em computadores pessoais. O programa apresentava a opção entre panoramas e objetos. Essa proposição pode ser vista até hoje, no método de captura de objetos e ambientes. Vídeos 360 que oferecem movimentos de cabeça são uma evolução deste conceito e uma expressão do extrínseco, em mímese de um olhar observador, proporcionando deslocamento, enquanto objetos em 3d são dispostos em realidade aumentada e virtual e podem ser vistos espacialmente. Esta divisão de característica aparece em dois projetos ligados à egiptologia, V-Horus, onde a narrativa se dá em torno da Múmia Romana e Neferhotep 360, onde a pessoa é levada para dentro da tumba tebana 49, em vídeo 360 documental.

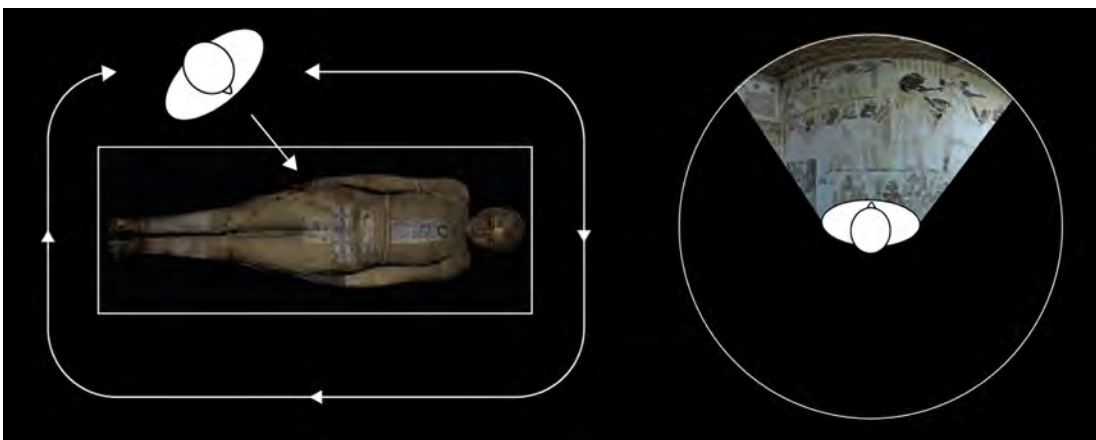


Figura 18 - Em uma vista superior a disposição de navegação intrínseca e extrínseca

Em espaços imersivos, ambos as direções coexistem, mas podemos notar que desenvolvimento de objetos e cenário muitas vezes também é separado tecnicamente. Um exemplo é que em metaversos por vezes, imagens equirretangulares são utilizadas como céu (*Skybox*), ambientes por sua vez, desenham as possibilidades de navegação. Objetos ou assets, são usados como itens, por vezes interativos. É possível notar que a geração de modelos tridimensionais por IA atende em sistemas diferentes para cada uma dessas categorias.

3.1 A câmera intrínseca

Uma das primeiras experiências do grupo Space-XR no metaverso foi a concepção da exposição virtual XR-Showcase³. Este espaço foi constituído para apresentar objetos e ideias referentes à estudos dos laboratórios Bio Design, da PUC-Rio, VIS-GRAF Lab do Instituto de Matemática Pura e Aplicada e LAPID do Museu Nacional do Rio de Janeiro, em um momento inicial de estudos em realidade virtual compartilhada. Foi utilizada a plataforma Spatial.io, na qual foi criado a partir de um modelo de espaço chamado Isle Gallery, um ambiente virtual desenhado para exposições. O Spatial é construído em Unity 3d⁴, e atualmente distribui um kit de desenvolvimento (SDK) que o torna mais personalizável, mas no momento da montagem do espaço as opções eram limitadas ao *upload* dos modelos. O fato de funcionar em diversos dispositivos, especialmente no HMD, e pela facilidade de acesso, foram os motivos para adoção da plataforma.

Aprender características deste método foi uma experiência interessante sobre ver objetos tridimensionais em uma experiência imersiva compartilhada, especialmente por possibilitar o diálogo entre os pesquisadores dentro do próprio aplicativo enquanto observavam os modelos tridimensionais. A dinâmica facilitada pela implementação

³ Disponível em <https://www.spatial.io/s/XR-Showcase>, acesso 20/12/25

⁴ Ver motores de jogos Unity 3d em <https://unity.com/> e <https://www.unrealengine.com/>, acesso 20/12/25

possibilitava um diálogo fluido diante dos modelos tridimensionais em realidade virtual e trazia uma maneira interdisciplinar de olhar para eles. Na montagem, os próprios autores participaram de uma experiência interdisciplinar de aspecto colaborativo (Mitchell; Novaes, 2024). A sensação na observação percebia a escala de maneira diferente, parecia buscar detalhes e verossimilhança com a física. Um paradoxo percebido pelo grupo da visualização do modelo em 6DoF, com a coerência de movimento de rotação da cabeça e deslocamento, é que existe um estímulo de aproximação ao objeto para observar detalhes enquanto os modelos tinham limites de resolução de textura e limite em número de polígonos.

Nesse caso, os objetos são fruto de captura e tratamento enquanto o cenário ajuda a organizar uma experiência ramificada (Mitchell; Novaes, 2024), dados os caminhos disponíveis, mas não se expressa muito enquanto forma, tal qual o desejo do cubo branco de uma galeria. Apesar do ambiente navegável, esse espaço virtual teve a exposição de objetos como principal objetivo.



Figura 19 - Sala Lapid no espaço XR-Showcase dentro da plataforma Spatial.io, disponível em: <https://www.spatial.io/s/XR-Showcase-614666928c2af60001d23190?share=2973599184567383091>

Para a sala LAPID deste ambiente foram tratados quatro objetos digitalizados do acervo do Museu Nacional para aplicação no espaço. Os modelos foram gerados através de três formas de captura distintas. O caixão de Sha-Amun-En-Su, por Pedro Seen Vonhausen, e o vaso Marajoara, representando um Felídeo, capturado por Jorge Lopes, foram capturados através de fotogrametria, enquanto o fóssil do crânio de Luzia foi tomografado por Jorge Lopes e o esqueleto do dinossauro *Staurikosaurus Pricei* foi escaneado e modelado por Orlando Grillo. Ao observar os objetos para implementação no espaço virtual, a equipe detectou algumas modificações que poderiam ser importantes, como a remodelagem em menor número de polígonos, a texturização e o sombreado, uma vez que em realidade virtual a representação (*shader*) utilizada pelo sistema não aplicava sombra sobre os modelos em realidade virtual. Conforme pôde ser avaliado empiricamente, a texturização e as sombras são tão importantes quanto a geometria para a percepção visual em um ambiente virtual.

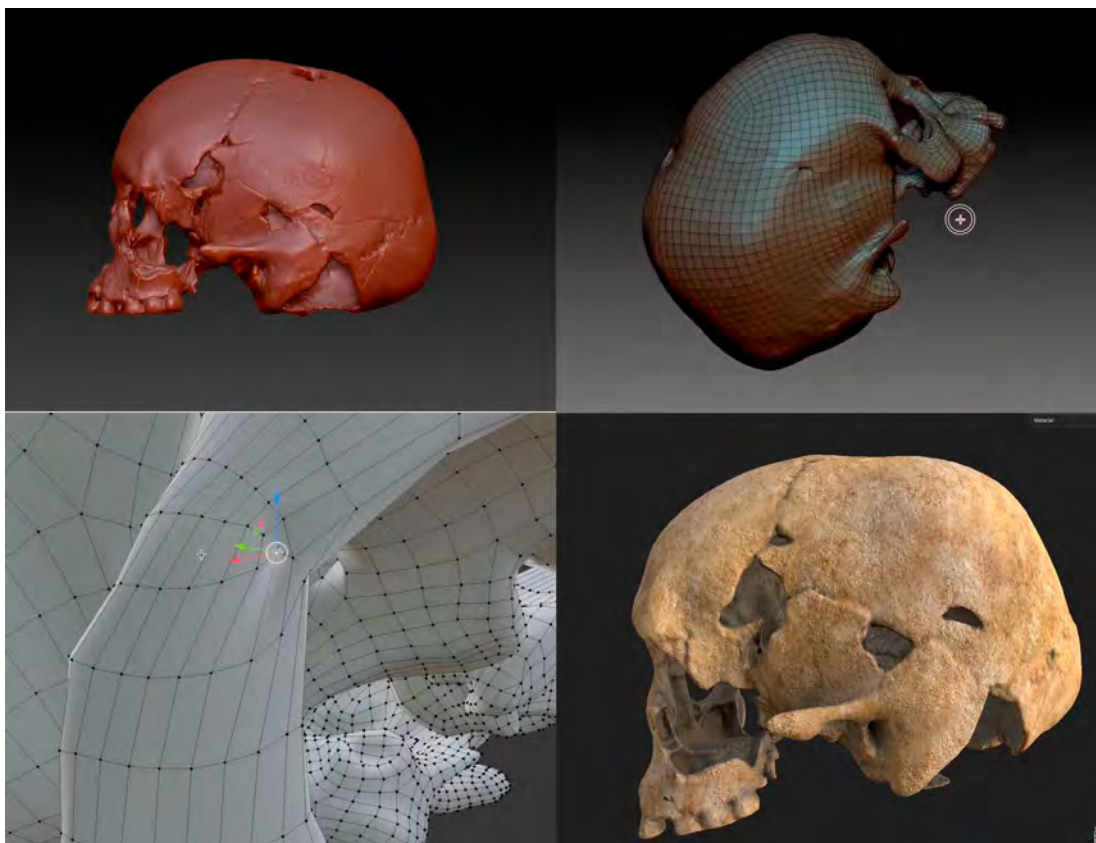


Figura 20 - processo de tratamento e textura do modelo do Fóssil da Luzia, da esquerda para direita o modelo bruto com milhões de polígonos, o segundo gerado através de remesh do zbrush, o terceiro em retoque, abaixo à esquerda, e o quarto com textura aplicada no adobe substance painter. Disponível no endereço acima

Para reduzir polígonos ⁵, foi feito o remesh no Maxon Zbrush, e para texturização e sombreamento, aplicado na textura (por *bake shadows*, que imprime a sombra na textura), Adobe Substance Painter, a avaliação e tratamento inicial e final sendo realizada no Blender. O caixão de Sha Amun em Su (“Caixão de Sha-amun-en-su - Egito Antigo | Museu Nacional - UFRJ”, [S.d.]) pouco foi alterado, mas como o autor deu acesso ao conjunto de imagens original, foi retreinado algumas vezes utilizando o Photoscan e Adobe Sampler. A busca por melhor amostragem era importante, uma vez que

⁵ Conforme tutorial realizado disponível em <https://www.youtube.com/watch?v=XvRhGW-CJbu8>

a textura poderia revelar as principais características do objeto. Era preciso que os desenhos e escritos fossem visíveis.

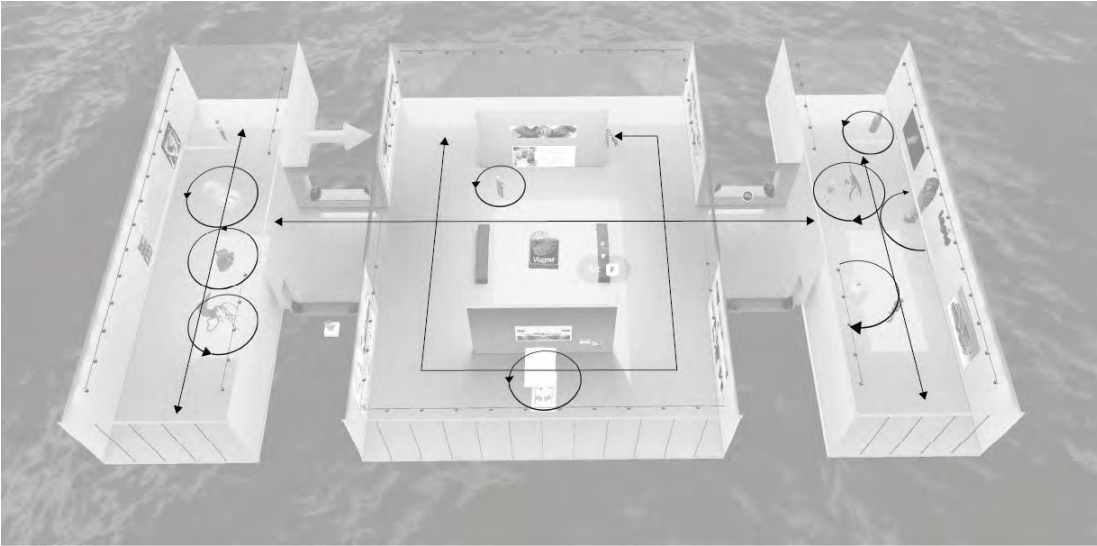


Figura 21 - Um estudo de navegação proposta pelo cenário e pelos objetos dispostos organicamente pelo grupo

Embora muitas imagens tridimensionais em 3DGS tenham sido treinadas, o trabalho em torno do caixão de Sha Amun em Su recebeu mais iterações. Sendo um objeto perdido, ter a melhor amostra tridimensional a partir dos dados existentes pareceu ser pertinente.

O caixão de Sha Amun em Su, possui importância histórica acumulada em várias camadas. Foi um artefato ofertado a Dom Pedro II e fez parte da coleção egípcia do Museu Nacional desde a proclamação da República (“Sha-Amun-en-su”, 2026). Formado por uma caixa e uma tampa, ambas talhadas em madeira estucada e policromada, tinha 1,58 metro de altura e foi produzido por volta de 750 a.C. Tratava-se de uma peça bastante representativa da arte funerária egípcia dos séculos VIII e IX a.C.

É um artefato que foi perdido no incêndio do Museu Nacional em 2018 e considerado raro, conforme disse o curador da coleção egípcia do Museu Nacional, Antonio Brancaglioni Junior, em 2015:

"Quem tem múmia tem, quem não tem não vai ter mais. Se a perdermos, jamais conseguiremos outra coisa remotamente parecida. Temos que conservá-la ao máximo" ("Sha-Amun-en-su", 2026)



Figura 22 - Comparação entre modelos gerados a partir de tomografia e fotogrametria e fotos do artefato

A busca por uma boa qualidade de amostra para o caixão de Sha Amun En Su foi encaminhada como teste para o novo formato 3DGS. Por outro lado, a configuração do objeto era boa para o modelo tridimensional em malha e foram obtidos resultados consistentes. Nestes testes foi gerado um modelo em malha com 1 milhão de polígonos e mapa difuso de 8k, para comparação em um plano ideal, em que a máxima qualidade fosse suportada.



Figura 23 - Modelo gerado em malha

Em 3DGS, o conjunto de fotos do caixão de Sha Amun En Su foi reprocessado muitas vezes, no primeiro momento usando o Luma Interactive, o que não gerou um modelo de qualidade, tampouco o ambiente, com o resultado bastante confuso, com muitos *floaters*⁶ depois, a partir do programa Epic Reality Scan, gerando coordenadas de posicionamento de câmeras e nuvem de pontos através do Colmap. Em seguida, foi processado no programa Lichtfeld Studio para treinar as gaussianas, esse processo foi feito algumas vezes para aprendizado. Embora seja um conjunto de imagens fotográficas em alta resolução, são um pouco mais esparsas do que seria ideal para a técnica, por outro lado o processo foi capaz de revelar detalhes das imagens presentes na superfície do caixão em qualidade comparável ao processo feito por malhas e texturas. O programa Jawsset Postshot também foi utilizado, porém apresentou um problema recorrente de transparência na superfície.

⁶ Nome dado a elipsóides (ou splats) residuais, por fluturarem



Figura 24 - O conjunto de imagens do caixão de Sha Amun em Su foi utilizado para o treinamento, ao treinar um pouco da sala onde foram feitas as imagens fica visível.

Uma vez que a geração inteira tem o resultado bastante ruidoso, foi testado o recurso de criação de máscaras para isolar o caixão do ambiente, dedicando o treinamento de splats unicamente ao caixão e especificando os limites. Para tal, foram testadas duas formas disponíveis no programa Epic Reality Scan, a primeira através da geração de um modelo 3D e a partir dele a impressão em imagens dos limites desse modelo. A outra forma é utilizando inteligência artificial, se assemelha bastante, mas na medida em que não usa o modelo, apenas a sugestão da forma, pode trabalhar com formas mais sutis.

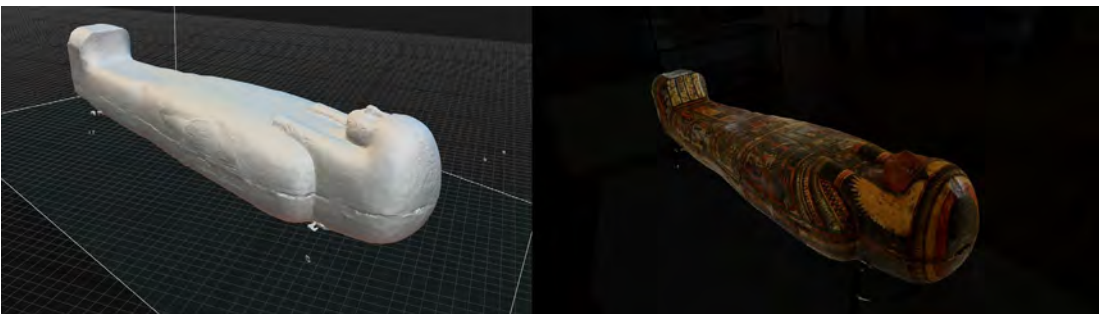


Figura 25 - processo de criação do modelo para geração da máscara

No Lichtfeld Studio foi utilizado um pré-tratamento com um programa adicional para densificar a nuvem de pontos (Edstedt et al., 2025). Para o treinamento, além da inclusão das máscaras, foi utilizado o recurso PPISP (Deutsch et al., 2026), que entende o comportamento da câmera através inteligência artificial, para então, interpretar as imagens, compensando luzes e calibrando-as. O treino então foi configurado até 20.000 iterações exportado e aberto no Supersplat para limpeza e ajustes.



Figura 26 - Modelo final em 3DGS, disponível para visualização em <https://superspl.at/scene/7d8feca>

Em paralelo, a título de comparação e aprendizado, foi treinado também o modelo da cartonagem de Irthorru em processo semelhante, fotografado por Daniel Pett, no British Museum (Pett, 2018). Apesar do processo ser quase idêntico, o conjunto de dados foi tomado de forma bem diferente. Enquanto no primeiro caso as imagens são da totalidade da forma, no segundo as imagens são em close-up. Os dois métodos funcionam e pode-se dizer que se complementam, uma vez que os detalhes são relevantes, nos dois casos, mas que também a forma total colabora com a coesão do modelo. Nesse

sentido, algumas diferenças aparecem coerentemente, como mesmo com o uso de máscara, o 3DGS e da cartonagem teve uma borda um pouco difusa, por outro lado os hieróglifos são muito nítidos. Desta forma, as duas abordagens são complementares.



Figura 27 - Imagem em 3DGS da cartonagem da múmia Irthorru, realizada através do conjunto de imagens de autoria de Daniel Pett, no British Museum, disponível em <https://superspl.at/scene/1431aac0>

No caso do caixão de Sha Amun En Su, comparando à uma versão gerada em malha com textura de 8k, o 3DGS ainda perde um pouco de nitidez em pequenos detalhes, por outro lado a reflexividade fotográfica do modelo supera uma sensação de iluminação real e apresenta alguns detalhes do relevo que o modelo em malha ignora.

Em seguida observamos o 3DGS sintético, este termo serve a 3DGS treinados a partir de modelos tridimensionais. A partir dos modelos do vaso marajoara e do fóssil de Luzia foram renderizadas imagens em cycles, opção do Blender que utiliza *raytracing*, para treino de gaussianas, a partir de programa adicional (*addon*), em um processo que é chamado treino sintético. Nesse caso funciona como um preparo ⁷, sedimentando

⁷ A tradução do termo *bake* pelo autor é de certa forma um preparo, é utilizado como conformar estaticamente características dinâmicas em objetos. Um bom exemplo são sombras que são impressas nos mapas de textura, ao invés estarem sendo produzidas pela exposição de uma luz. Esse tipo de preparo é amplamente utilizado para economizar o trabalho da placa gráfica.

as características de qualidade alta, sombras e nuances vindas do modelo tridimensional em alta complexidade nas gaussianas. No caso de Luzia foi testado também um sombreamento transparente.



Figura 28 - Modelo 3DGS sintético feito a partir do Blender utilizando shader de vidro e uma textura de ossos disponível em: <https://superspl.at/scene/a0d75ae3>

Destas montagens surge a percepção de que o 3DGS pode ser não apenas um formato de captura, mas um potencial formato de distribuição de imagens, isto é reforçado ao observar a galeria supersplat com diversos exemplos de 3DGS sintéticos, ou seja, a disposição de imagens em campo de radiância pode vir a ser um novo modo de ver imagens em volume independente do modo de captura, incluindo imagens geradas por 3D, IA ou qualquer processo volumétrico de imagens.

3.2 A câmera extrínseca

Talvez uma das vocações da realidade virtual, o deslocamento, pode ser obtido de várias maneiras. Cinema, literatura e outras artes possuem propriedades imersivas,

a depender de suas linguagens, mas também das obras e suas características. A realidade expandida parte da premissa que o deslocamento pode ser físico-visual, em sistemas chamados 3DoF e 6DoF, que significam três graus de liberdade ou seis graus de liberdade. O primeiro se refere aos movimentos de rotação cabeça, para cima/baixo, giro e inclinação e o segundo aos três eixos de deslocamento X,Y,Z. Espaços baseados em malhas e 3DGS são, em geral, 6DoF. Filmes 360 graus são 3DoF, por outro lado é o “olhar em torno” ou ao redor, logo, o movimento mais importante para compreensão de ambiente.



Figura 29 - O nicho com as esculturas de Neferhotep e sua esposa é a expressão iconográfica do sítio

Diante da possibilidade de pensar a criação em novas mídias, o filme Neferhotep 360 foi proposto por Luiz Velho para o grupo Space-XR e realizado a partir de imagens tomadas por Pedro Seen Von Seehausen acerca do trabalho da equipe de pesquisa em egiptologia sobre a tumba tebana 49, ou seja, a tumba de Neferhotep, “um nobre egípcio do tempo do Novo Império, aproximadamente 3.300 anos”, como descreve o texto narrado por Sergio Azevedo no filme. Os dois temas, a pesquisa do grupo de egiptólogos e a história da tumba são entrelaçadas conforme as informações das pesquisas são

reveladas. Os pesquisadores, em especial Professora María Violeta Pereyra, são personagens apresentados. A imagem do ser humano em VR, nesse caso, parece reforçar a percepção documental da experiência, porque passa a ter ações registradas.



Figura 30 - Professora María Violeta Pereyra na TT49

O filme foi editado como parte desta pesquisa em parceria com Luiz Velho e foram utilizados dados como a fotogrametria, a geolocalização e as próprias imagens editadas em Adobe Premiere e trabalhadas graficamente em Adobe After Effects. Outra característica foi a manipulação da percepção de proximidade da Tumba, onde a projeção das imagens foi alterada para intensificar a sensação do espaço em torno e favorecer a percepção de deslocamento. Para isso, foi utilizado um plugin de Adobe Premiere chamado Mantra da Mettle. Uma cadência lenta e contemplativa foi escolhida para que o observador fosse conduzido para perceber os detalhes das imagens.

As imagens 360 possuem índice do real e durante a edição foram utilizados modelos tridimensionais da pesquisa para ilustrar a descrição. Esse índice real não existia em ambientes de metaverso, mesmo em ambientes que utilizam sombras e modelos em

fotogrametria. Talvez o dado de realidade registrado em pixels indique uma nuance do sentimento de presença que a tridimensionalidade não possui.

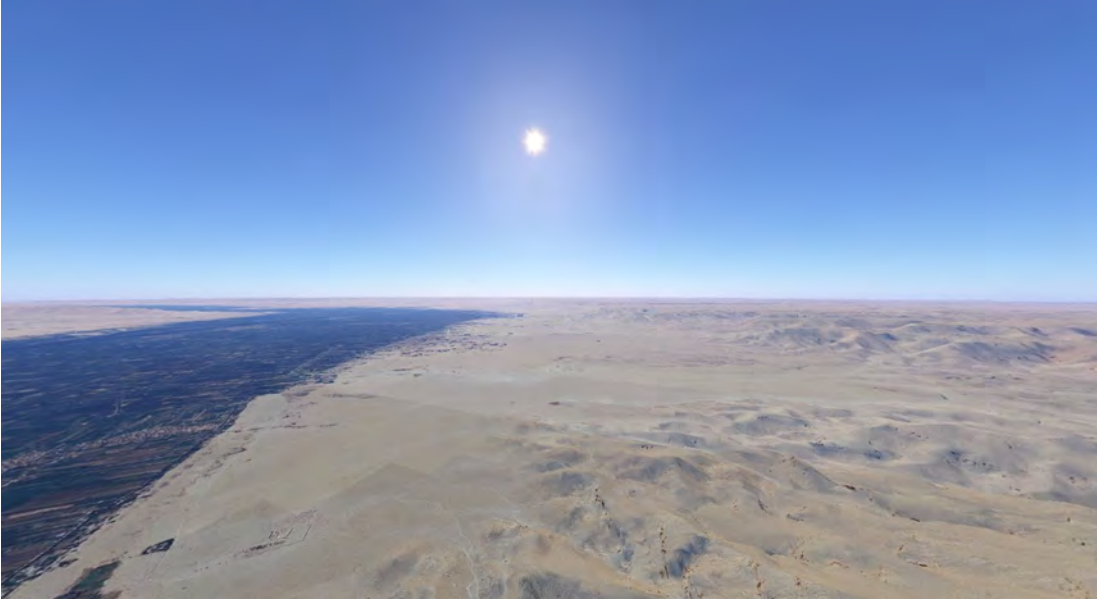


Figura 31 - Imagens geolocalizadas geradas pelo Google Earth Studio

A edição com outros dados, como geolocalização, demonstrada através de imagens extraída do Google Earth Studio e fotogrametrias realizadas por Pedro Seenvonhausen, propõem aumentar a percepção de informação a respeito do sítio de pesquisa. O videografismo em uma das paredes tem o propósito de guiar o olhar e orientar o espectador a observar as áreas comentadas.



Figura 32 - O trabalho de videografismo foi realizado para auxiliar a compreensão entre imagens e locução

Em sua construção, a edição do filme traz algumas reflexões em imagem sobre a virtualidade presentes no estudo, através de fotogrametria e as imagens em vídeo 360. O uso de ferramentas de videografismo e tratamento de imagem apresenta uma ponte entre o audiovisual e as experiências imersivas e cria a possibilidade de utilização de ferramentas de design em movimento.

As imagens apresentadas podem ser desdobrar em várias perspectivas: educacional, entretenimento e documental. Para a contextualização e sensibilização, as imagens filmadas são importantes pois comunicam a possibilidade do real. Essa imagem é enriquecida quando é seguida por áudio referente aos estudos da equipe sobre a tumba e seus adornos, principalmente quando as imagens são adequadas a visualização espacial da mídia, como a fotogrametria do nicho com a escultura de Neferhotep e Merytha, sua esposa.

O nome foi discutido entre a equipe e algumas premissas apareceram, em um primeiro momento a necessidade de apresentar o título Neferhotep como nome direto foi aceita, mas para especificar ao conteúdo científico, foi acrescido um subtítulo Teban Tomb 49, com um comentário válido que existiram outros Neferhotep e que a identificação do sítio estudado era essa. Ao título principal foi acrescida a especificação da forma do projeto 360, o que também colabora com a expectativa. Foi criado então uma

forma do título, utilizando Adobe Illustrator para desenho. Foi animado e conformado em imagem equirretangular no Blender, para ser inserido junto às sequências editadas no Adobe Premiere. Além disso, foi aplicado em cartazes, banners e camisetas. A solução tipográfica buscou estabelecer uma ponte entre estudos clássicos e uma aventura, utilizando ligaturas e contrapondo com uma disposição contemporânea. A utilização de texturas e da luz, na aplicação do filme, procura criar expectativa, mistério e vínculo ao interior da tumba. A utilização do tipo serifado busca evidenciar essas características, juntamente com certo rigor no estudo. Evitamos que a este título, a identidade do filme em sua forma, se parecesse com tipos egípcios, justamente para não transmitir uma imagem estereotipada conforme discussões recentes (“Stereotypography”, 2020), algo que contrastado ao estudo propriamente poderia soar caricato.



Figura 33 - Vinheta de abertura com o design do título

O fato de as imagens serem editadas, sincronizadas a uma locução descritiva com trilha sonora, cria sentido para a experiência em sua visualização. Se fossem apresentadas soltas, sem uma narrativa, as imagens não engajariam um público mais amplo, como foi o caso, um público que foi este composto desde crianças até estudiosos no

assunto. O trabalho foi exibido no Festival da Matemática 2024, na Semana de Egip-
tologia do Museu Nacional do Rio de Janeiro 2024, no Museu do Cairo e nas Mostra
de Realidade Virtual na Semana Nacional de Ciência e Tecnologia.

Em metaversos como o Spatial.io, o grupo Space-XR experimentou algumas pos-
sibilidades de ambientes, tanto construídos quanto inteiros, em fotogrametria ou mis-
tos. Uma equação difícil é a relação entre extensões maiores e a dificuldade em manter
qualidade, principalmente naquele momento específico, pois atualmente as possibili-
dades se alargaram um pouco.

Um aspecto a ser observado é que a visão em primeira pessoa em realidade virtual
é exigente em relação a detalhes de representação física, porque quer imputar a reali-
dade no ambiente, uma vez que se move realisticamente.



Figura 34 - Metaverso criado a partir de fotogrametria

Alguns ambientes nesta pesquisa foram criados referentes a modelos *open-
source* disponíveis no Sketchfab, como La Motilla del Azuer (“La Motilla del Azuer -
Daimiel, Spain - Download Free 3D model by Global Digital Heritage and GDH-
Afrika (@GlobalDigitalHeritage)”, 2018). Este metaverso foi utilizado como um teste
de percepção de presença, cabe ressaltar que o mesmo sítio foi objeto de um trabalho

para realidade virtual, porém uma característica de metaverso se aplica neste experimento: a disponibilidade online. Ambientes inteiros de fotogrametria, como o Tartaruguito (“Spatial - Create Immersive UGC, Virtual Classrooms, Experiential Marketing”, [S.d.]) realizado por Orlando Grillo, funcionaram bem em espaços menores, assim como soluções mistas como presentes nos espaços do LAPID. Cabe ressaltar que o Spatial.io limita bastante o tamanho, peso e complexidade das malhas dos arquivos e resolução dos mapas de textura. Quando isso é levado para ambientes torna o desenvolvimento a partir de fotogrametrias bastante complexo.

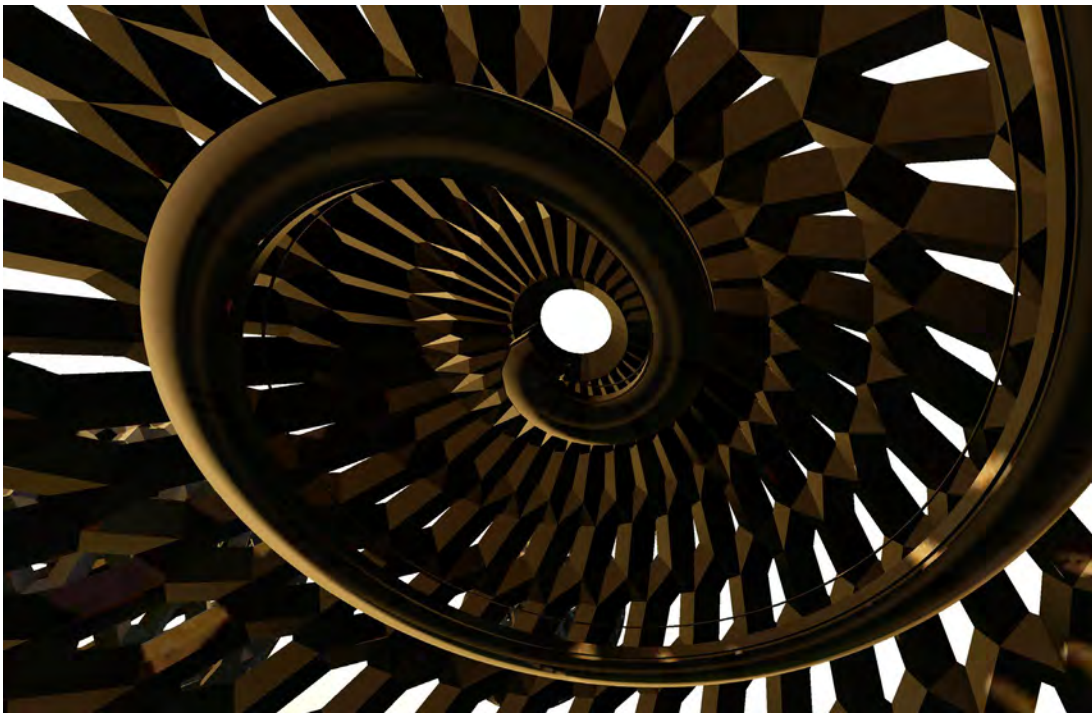


Figura 35 - Experimento de ambiente e navegação

Em modelagem os cenários apresentam muita liberdade, especialmente no Spatial.io após a disponibilização do SDK, pacote de desenvolvimento em Unity3D. Na medida em que são elaborados, servem de guia para navegação no espaço, sendo indicados para apresentação de objetos e outras aplicações. É interessante pensar que ao criar um espaço virtual, em programas como Blender, todo o pensamento referente a

identidade visual deste espaço pode ser aplicado. Aspectos como paleta de cores, tipo de luz e características da forma podem ser controlados para uma experiência estética coerente com um projeto. Foram feitos diversos experimentos nesse sentido que nos levam a aprender sobre sensações físicas em espaços virtuais, sombras, texturas, usabilidade e navegabilidade.

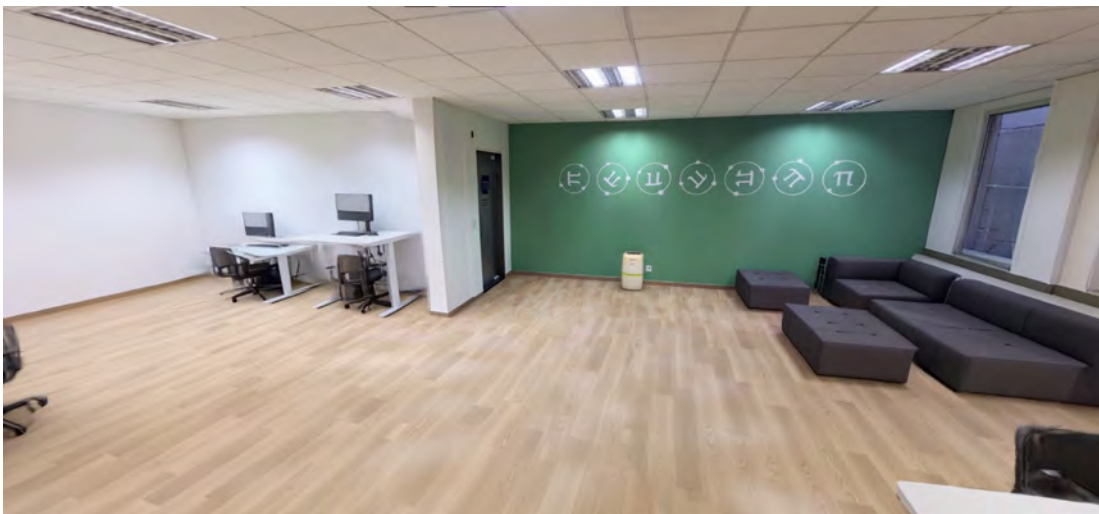


Figura 36 - Sala do Centro Pi em 3DGS

A primeira experiência de reprodução de ambientes do grupo, a sala Centro Pi, foi modelada nesta pesquisa em parceria com Vinicius Arcoverde e Mariana Duarte e

reinterpretada através do sistema Marble/Wordlabs, sistema que utiliza inteligência artificial para produzir cenários em 3DGS. Embora tenha várias inconsistências e não tenha muita qualidade no modelo, já que o Marble interpreta a referência, a comparação se faz útil entre a modelagem em precisão geométrica e a estimativa.

Levamos a expectativa de deslocamento 3DGS a captura de ambientes, observando a possibilidade de utilizar a aparência fotográfica. Alguns projetos expostos na plataforma Supersplat impressionam, como por exemplo a garagem de Simon Bethke, que criou um tutorial de como alcançar resultados bastante precisos (Simon Bethke, 2024), incluindo o uso de APRIL tags e uma estratégia de captura espacial baseada em voxelização do espaço e captura adicional de detalhes.

Antes de conhecer esse guia, foi utilizado um pequeno ambiente do Jardim Botânico do Rio de Janeiro, a estufa de insetívoras, além de outros ambientes do próprio Jardim Botânico como espaço de experimento de captura.



Figura 37 - Experimento de captura de ambiente

Neste mesmo espaço foi testada a captura utilizando uma câmera 360. Câmeras 360 têm se tornado um instrumento de captura de ambientes com a qualidade de maior

velocidade, visto que captura vários ângulos simultaneamente. Em geral a imagem capturada é fatiada através de programas como 360 Extractor⁸ (“360 Extractor tool for Blender”, [S.d.]), para depois passar pelo processo *de surface from motion* no Epic Reality Scan e posteriormente treinar gaussianas. Na medida em que o momento da captura é o mais crítico, e são necessários múltiplos pontos de vista, algumas pessoas criam rigs, estruturas com múltiplas câmeras, como o próprio Olli Huttunen. Por outro lado, a qualidade ótica dessas câmeras, em geral, é bastante inferior às DSLR ou mirrorless. No entanto, mesmo nessas devem ser usadas lentes grandes angulares, pois a área de captação mais larga diminui significativamente o esforço de captura. Existem soluções para eliminar a distorção ótica resultante das lentes como 3D GUT (Wu et al., 2025).



Figura 38 - Experimento de captura de ambiente

A ideia de múltiplas câmeras para cenários foi incorporada pela Meta na criação de um aplicativo chamado Hyperscape, o aplicativo faz uso das diversas câmeras pre-

⁸ Existem vários programas surgindo, de pequenos desenvolvedores, como Olli Huttunen, autor deste. Existem também um fluxo utilizando Metashape em

sentes nos HMDs Metaquest 3 e Pro para captura de ambientes. Eficaz na representação, o aplicativo peca por não disponibilizar o arquivo resultante para download, sendo visível apenas no próprio aplicativo. Partindo da mesma concepção, existem aplicativos opensource em desenvolvimento, como o Splatara. Na mesma perspectiva de multivisão, a XGrids, uma empresa de scanners 3D, criou um equipamento específico para a captura de ambientes em 3DGS chamado Portal Cam (“XGRIDS - 3D Spatial Intelligence Technology”, [S.d.]), com sensor Lidar e câmeras fisheye integrados.

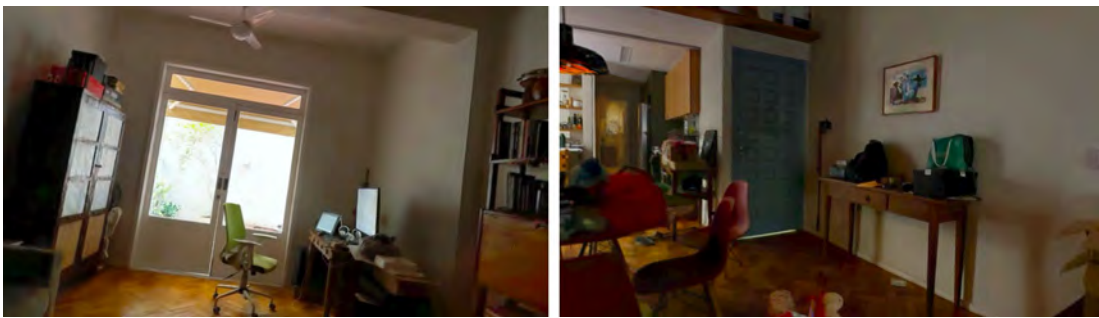


Figura 39 - Captura e visualização através do Hyperscape da Meta

3.3 Distribuição

A forma da imagem tridimensional e o planejamento de sua visualização pode tornar mais aplicável à realidade virtual ou realidade aumentada. A forma em si não determina, ambientes podem ser apresentados como maquetes, objetos podem ter aplicações imersivas, então uma camada de planejamento de como será a visualização pode dirigir a intenção sobre a experiência a ser proporcionada e preparar as condições para uma experiência.

São diversas formas de apresentação de conteúdos tridimensionais no ambiente digital, experiências ricas em multimeios ou a disponibilização imersiva direta, a possibilidade de ver em conjunto com outras pessoas ou em experiência individual. Aqui incluo algumas formas diferentes de disponibilização de arquivos tridimensionais para experiências, tanto através do computador quanto headsets e celulares. Assim, existem

variáveis como a responsividade entre dispositivos, compartilhamento em tempo real e utilização das propriedades de hardware, como visualização imersiva, câmeras e potência de processamento.

Nesta pesquisa podemos incluir parte da experiência da digitalização do livro Realidade Expandida - Novas Mídias e IA (Azevedo et al., 2025) disponibilizado para um formato online ⁹. O site utilizou o wordpress, um sistema de manejo de conteúdo de código aberto, como indexador dos conteúdos. Essa mudança de suporte utilizou o benefício do amplo suporte de mídias e foi elaborado com conteúdo de modelos, metaversos e filmes. Durante o processo, observamos os modelos tridimensionais e os metaversos no contexto da divulgação científica e em sua diferença de forma de distribuição. Uma referência foi o *website* do Metropolitan Museum of Art, na seção de obras digitalizados tridimensionalmente (“Metropolitan Museum of Art”, [S.d.]

Em relação aos modelos, foi utilizado plug-in wordpress baseado em um sistema chamado Model Viewer, criado pela Google, que facilita a implementação de modelos principalmente utilizando o formato glb, convencionado como um formato tridimensional de malhas para a utilização online. Este modo possibilita que cada modelo seja visto na própria página em 3D ou em realidade virtual ou aumentada, tanto utilizando computadores, *smartphones* ou HMD. Nesse caso é interessante notar as diferenças entre dispositivos, enquanto o primeiro obtém uma maior qualidade gráfica, o segundo e o terceiro viabilizam realidade aumentada, móvel em smartphones e imersiva em HMD e a realidade virtual propriamente dita funciona apenas no HMD.

Essa perspectiva remete a responsividade, uma prática na construção de interfaces que gira em torno de sua adaptação em diferentes dispositivos. Além da adaptação visual, as funcionalidades dispostas e a experiência apresentam diferenças significativas. Quando indexamos um conteúdo tridimensional diretamente em um website, ele pode ser visualizado pelo HMD, mas sua contextualização está na página web em que

⁹ Disponível em <https://visgraflab.impa.br/xr-ai/>

está incorporado. Traz ferramentas para visualização em realidade virtual ou aumentada, mas a navegação é ligada ao website. A utilização de AR, ou mesmo a dinâmica de ver um objeto por vez é menos imersiva por não contar com outros recursos, mas é escalável e acessível.



Figura 40 - Exemplos de utilização em realidade aumentada

Aqui é interessante observar que para a apresentação dos objetos a realidade aumentada pode ter um interesse especial, onde se experimenta a virtualidade sem desconectar a visualização do ambiente, onde a experiência tridimensional se mistura ao acaso localizado pelo usuário. Em metaversos e outros tipos de experiência, como vídeo 360, a imersão é uma desconexão do entorno, o que pode aprofundar a experiência.

Essa popularização do arquivo tridimensional através da internet remonta ao VRML (Virtual Reality Modeling Language (Reis, 1998), uma das primeiras formas

de visualização tridimensional disponíveis na internet e fala sobre o desejo de utilização tridimensional pelo browser como uma alternativa ou complemento a visualização bidimensional dominante. Podemos avaliar que este desejo se relaciona com aspectos técnicos, visualizar e manipular arquivos tridimensionais e ambientes, assim como e recreativos, jogar, fruir conteúdo escultórico etc. Diversas formas de visualização para web surgiram até chegarmos ao WebGL e webXR.

Em projetos de metaversos onde se aplica uma experiência mais longa é necessário planejamento para integrar navegação, conteúdo e interatividade. A intensidade da experiência pode ser potencializada por recursos audiovisuais e navegabilidade. Este formato possibilita um primeiro acesso à realidade expandida a um público amplo, visto que é conectado e responsivo e se adapta a cada dispositivo.



Figura 41 - Metaverso criado a partir do projeto V-Horus

Em instalações com equipamento e estrutura física em experiências como V-Horus (“V-Horus – Virtual Egyptology”, [S.d.]), projeto que apresenta a múmia romana ¹⁰

¹⁰ Conhecida como múmia romana pois data do século I dC. época da dominação romana sobre o Egito.

em realidade virtual, são construídos diretamente nos motores de jogos, como Unity 3d e Unreal. Este utiliza narração do egiptólogo e curador, trilha sonora e iluminação dinâmica, mas é um ambiente de apenas um usuário. Este sistema funcionou em acesso direto ao público, fornecendo equipamento para experiências locais. A experiência V-Horus fez um extenso uso da iluminação em sua narrativa, o que é difícil de ser implementado sem o uso da placa de vídeo local. Foi feito um metaverso (Figura 39) com o modelo e características do projeto acessível através da web. Se por um lado pode ser visto junto com outras pessoas no metaverso, não há dúvida de que a profundidade da experiência é com a iluminação dinâmica e áudio sincronizado. Além disso, fez uso de elementos físicos, uma mesa, no caso relacionando real e virtual. Protótipos e projetos em novas tecnologias tendem a necessitar de estruturas específicas. Dadas essas diferenças, é possível pensar as características dos recursos disponíveis para visualização, desde a construção de sistemas até a utilização de plataformas.

Os metaversos ocupam nichos há tempos, com Second Life, Roblox e foram uma promessa tecnológica. Durante a pandemia de covid tiveram seu uso ampliado pela convergência entre distanciamento social e disponibilização de HMDs acessíveis. Esperava-se que este fosse um ponto de expansão (Baszucki, 2021), mas não foi, após a pandemia o que se viu o retorno ao uso em nicho, como lazer, estudo e uso profissional para visualização tridimensional (Kobie, 2025). Utilizar o HMD como recurso de visualização faz parte das funções iniciais do dispositivo, essa visualização é potencializada quando coletiva, porém o que se esvai, porém é a expectativa de acesso generalizado.

Para implementar 3DGS no sentido de metaverso surge o Arrival Space (Arrival.Space, [S.d.]), que propõe a possibilidade de criar espaços virtuais especialmente utilizando 3DGS, mas também modelos em glb, áudio e recursos de texto, em plataforma social. Também vale notar a viabilização de 3DGS no VRChat, uma grande plataforma social em VR (ou um metaverso). Esta plataforma possui a possibilidade de usar modelos em malha e 3DGS, então é possível pensar que ambientes virtuais compartilhados são uma forma de apresentação que atende a determinados requisitos.

Na perspectiva de compartilhamento de visualização de simples modelos como Sketchfab foi criado o Supersplat, que conta ainda com um editor de splats integrado, por ser *open source*, possibilitando a implementação em hospedagens independentes. Observando as aplicações, incorporar em um website cria um acesso de ampla distribuição através da internet, indexável aos buscadores e com sua consulta sob demanda, disponível a todo tempo. Por outro lado, embora viabilize a experiência tridimensional, não explora recursos de áudio que podem ser complementares e apenas parte das pessoas irá se dispor a fruir aquele conteúdo em modo imersivo.

A presença de narrativas, sejam em vídeos 360 ou metaversos, torna a experiência significativa (Mitchell; Novaes, 2024), enquanto a disponibilização individual de cada modelo supre uma demanda técnica especializada, mas pode também democratizar o acesso. Um projeto como Neferhotep 360, quando exposto no Youtube, terá uma visualização diferente de quando exposto em um evento, desde seu acesso até interface em torno. Do ponto de vista numérico, o acesso simplificado pode impressionar, uma experiência como a *Ocean Dance*, que teve aproximadamente quatrocentos e cinquenta acessos, somando sua apresentação no Seminário Danc3 & T3ch aos números de acesso em sua versão editada em vídeo disponível no Youtube, obteve vinte sete mil acessos, mas é uma Machinima do projeto original, onde embora a dança e a computação gráfica estejam representadas, não há a visão interativa.

Em outro caso, mantendo as características do arquivo, é possível esperar uma grande diferença na experiência de um filme 360, como Neferhotep 360 em um evento, com as condições preparadas para experiência e a grande maioria das pessoas que possam acessá-lo no Youtube. Ocorre de forma semelhante à mudança da experiência cinematográfica quando transposta para o streaming de filmes.

Se por um lado a internet amplia o acesso e faz com que a conectividade seja ubíqua, para instalações em eventos é mais prático e direto usar arquivos instalados no HMD ou no computador que provê acessos, evitando oscilações e eventuais esperas. O que se percebe é que a transposição da tridimensionalidade para diferentes formas de

implementação muda a forma de acesso e a experiência, pois as narrativas entre metaversos, instalações e websites, possibilitam diferentes recursos e sensibilizam de maneira diversa.

4.

Inteligência Artificial como ferramenta para expandir a imagens

Essa pesquisa trabalha muitas vezes em parceria com objetos e conjuntos de imagens que são frutos de pesquisas do Museu Nacional do Rio de Janeiro, instituição que sofreu um incêndio em 2018. Para criar experiências, precisamos completar lacunas de imagens ou conectar imagens. Nessa avaliação, apesar de considerar o 3dgs, parece mais construtivo observar consistência, características e recursos do que se ater ao formato, especialmente diante da possibilidade de passagem entre modelos 3d e 3dgs.

Diversos modelos de geração de objetos e imagens bidimensionais e tridimensionais e de vídeo foram testados, com diferentes características, mas faremos uma síntese documentada exemplificando. Desta forma, avaliamos de maneira qualitativa a possibilidade de criar imagens e volumes a partir de imagens utilizando IA. Faz-se necessário ressaltar que não cabe aqui observar toda a conjuntura do impacto da nova tecnologia, mas observar sob a perspectiva prática em exercícios, possibilidades de uso na criação de experiências.

4.1 Conectando e abrindo caixas

Podemos comparar o processo de treinamento do campo de radiância como uma revelação fotográfica, desse modo parece interessante apresentar a imagem da caixa preta, discussão proposta por Flusser sobre a dinâmica da fotográfica analógica. Na teoria de Vilém Flusser, a caixa preta representa um aparato técnico cuja lógica interna

permanece oculta ao usuário. Embora Flusser utilize a câmera fotográfica como exemplo central, sua teoria fundamenta a crítica política e ética ao software proprietário e valida o *open source software*, isto é, programas de código aberto.

Uma "caixa preta" é um sistema onde apenas a entrada (input) e a saída (output) são visíveis; os processos internos — o "programa" — são desconhecidos. O usuário de tecnologias de caixa preta torna-se um "funcionário" do aparato. Ele opera dentro das possibilidades pré-definidas pelo programa, em vez de exercer liberdade criativa real, assim como um poder reside no programador que projeta a caixa, pois é ele quem determina o conjunto de ações possíveis para o usuário. Os movimentos de software livre e de código aberto buscam "abrir" a caixa preta descrita por Flusser.

Ao publicar o código, os pesquisadores do INRIA (Institut National de Recherche en Sciences et Technologies du Numérique) não apenas entregaram uma imagem técnica (o resultado 3D), mas entregaram o programa inteiro. Isso permitiu que a comunidade global visse as "engrenagens" matemáticas (as equações de covariância e transparência) e implementassem inúmeras variações, programas e derivados. De maneira parecida o acesso ao sistema de IA generativa, Stable Diffusion (Rombach et al., 2022) facilitou o desenvolvimento de toda uma rede *open source*, abrindo a caixa. A partir do Stable Diffusion todo um ecossistema *open source* foi construído, o que viabilizou acesso de milhões de pessoas a tecnologia emergente.

A caixa preta da própria fotografia é também aberta através de um uso não programado do aparelho fotográfico. A câmera é usada como um scanner em configuração diversa da tradicional, enquadramento, velocidade, abertura, todos os parâmetros são alterados em função de critérios diferentes para gerar um 3DGS. É natural pensar nestes processos como caixas conectadas. A metáfora de caixa também interessa ao ser aplicada aos processos e decompor cada etapa em várias caixas, configuráveis.

Usar diagramas como interface para manipulação tem origem em programas como Pure Data para tratar do áudio e se populariza como programas de performance ao vivo como Touchdesigner, VVVV e Troikatronix Isadora. Posteriormente, foi ado-

tado ComfyUI, Fauna Flora e Weavy, e parece ser uma nova configuração para geração visual em que são cruzados sistemas, informações e imagens. Enquanto os primeiros foram pioneiros tanto em produção de imagens generativas quanto em interfaces de nós, os que seguem são interfaces de IA em lógica sofisticada. Mesmo para apresentar processos, as caixas parecem ser pertinentes, visto que tornam visual e interativa a montagem de fluxos de geração.



Figura 42 - Interface de "caixas" ou nós do ComfyUI

A cada caixa, uma entrada ou transformação dos dados, embora pareça hermética, dependendo da forma de construção, do sistema, pode ser acessível. Dessa forma, é possível transformar os sistemas, por vezes, com simples alterações ou mudando seus dados de leitura. A metáfora das caixas pode ser visível como nos programas citados acima, ou invisível, mas presente quando concatenamos processos nos quais podemos conduzir e tensionar os resultados de forma consciente ou experimental.

Neste contexto exploratório, realizamos um estudo experimental sobre o uso de inteligência artificial generativa na produção de imagens a partir de um conjunto de ilustrações autorais. O experimento partiu de uma série de desenhos produzidos para os vídeos do projeto *Capoeira no Rio de Janeiro 1948-82*, vinculado ao site *Capoeira History*. As ilustrações originais foram realizadas manualmente em meio digital, com

o objetivo de compor cartelas e vinhetas animadas, tomando como referência imagens dos personagens históricos abordados nos vídeos. Visualmente, a série foi construída de modo a remeter ao mimeógrafo, em diálogo com técnicas de reprodução gráfica associadas ao período tratado.



Figura 43 – Exemplos de imagens utilizadas no treinamento

O objetivo do experimento não foi substituir a prática do desenho, mas investigar de que modo sistemas generativos poderiam operar a partir de um repertório visual já existente, produzindo derivações imagéticas associadas a ele. A questão central consistiu em observar se seria possível treinar modelos para gerar imagens próximas de um determinado estilo autoral, e em que medida esse processo poderia contribuir para reflexões sobre design, ilustração, autoria, viés e cultura visual. Nesse sentido, o estudo adotou uma abordagem prática e exploratória, entendendo o experimento como procedimento de investigação. A caixa mais opaca é o espaço latente, o modelo treinado, do qual é extraída qualquer imagem de IA generativa.

O trabalho dialogou com experiências anteriores de geração por inteligência artificial a partir de acervos e repertórios artísticos, como *The Next Rembrandt* (Microsoft, 2016), bem como com pesquisas que discutiram o uso de coleções museológicas no treinamento de sistemas generativos. No entanto, nesse caso, o interesse esteve menos na reconstrução de obras existentes e mais na possibilidade de o próprio ilustrador utilizar o treinamento como recurso de criação e experimentação. O ponto de partida foi a hipótese de coexistência entre desenho e inteligência artificial, considerando que sistemas generativos poderiam ser incorporados ao processo projetual sem eliminar a dimensão autoral do trabalho humano.

Do ponto de vista técnico, o experimento utilizou o modelo Stable Diffusion, escolhido principalmente por sua licença aberta e pela possibilidade de treinamento adicional por meio de métodos como DreamBooth e LoRA. Essa abertura permitiu testar não apenas a geração de imagens, mas também formas de adaptação do modelo a um conjunto específico de ilustrações. Ao mesmo tempo, o estudo reconheceu que esse tipo de sistema foi treinado sobre grandes bases de dados coletadas da internet, como LAION, o que implicou limitações importantes relacionadas a vieses culturais, raciais, etários e de representação. Assim como entre os dois estágios, existe uma mudança de versão, passando do Stable Diffusion 1.5 para o SDXL. Assim, a investigação técnica esteve articulada a uma reflexão crítica sobre os condicionamentos inscritos nos modelos.

Na primeira etapa, foi utilizado o método DreamBooth. As ilustrações foram reorganizadas para o treinamento: retiraram-se letreiros, realizaram-se cortes em formato quadrado e as imagens foram redimensionadas para a resolução exigida pela plataforma. O treinamento ocorreu a partir do site getimg.ai, com a definição de uma palavra-chave destinada a ativar o estilo aprendido pelo modelo. Os resultados mostraram que o sistema absorveu algumas características importantes da série original, como enquadramentos de retrato, traços desenhados sobre base fotográfica e uso de contrastes fortes nas sombras. No entanto, a coesão do estilo permaneceu parcial. As cores apareceram instáveis, certas soluções visuais foram alteradas e o modelo produziu imagens

que misturaram traços do conjunto original com padrões já presentes no modelo de base.

Apesar dessas limitações, o treinamento revelou um aspecto importante: o estilo aprendido não dizia respeito apenas à aparência gráfica, mas também ao tipo de figura humana predominante nas imagens originais. Como o conjunto de treinamento era composto majoritariamente por retratos de homens negros e idosos ligados ao universo da capoeira, as gerações passaram a reproduzir também essa orientação representacional, inclusive quando o prompt utilizado era genérico. Em vez de repetir o viés mais comum dos modelos amplamente treinados na internet — que tenderam a privilegiar representações brancas, jovens e ocidentalizadas —, o modelo treinado apresentou uma espécie de inversão, gerando frequentemente homens e mulheres afrodescendentes, muitas vezes idosos. Esse resultado tornou visível a relação entre estilo e representação, mostrando que a aprendizagem do sistema não se limitou ao traço, mas incorporou também padrões de corporeidade, idade e etnia.

Para testar esse comportamento, foram utilizados prompts genéricos como “man with hat” e “portrait of woman”, comparando os resultados do modelo treinado com outros modelos voltados a universos gráficos distintos. A comparação evidenciou que diferentes modelos responderam de maneiras diferentes ao mesmo comando, conforme os dados e referências visuais que estruturaram seu treinamento. Isso reforçou a ideia de que os vieses presentes nas imagens geradas não foram casuais, mas resultaram das bases de dados, das classificações e das culturas visuais dominantes que alimentaram esses sistemas. No caso do experimento, o treinamento com imagens ligadas à capoeira e a personagens específicos deslocou parcialmente esses padrões dominantes.



Figura 44 - Primeiros resultados da experiência

Ainda assim, alguns elementos importantes das ilustrações originais não foram bem assimilados. Componentes arquitetônicos quase não apareceram nas gerações, provavelmente porque ocuparam posição secundária no conjunto de treinamento. O berimbau, instrumento central na cultura da capoeira e presente em várias imagens originais, surgiu poucas vezes e frequentemente deformado. Esses problemas indicaram limites do treinamento quando certos elementos apareceram em menor quantidade ou quando a visão computacional não os reconheceu adequadamente.



Figura 45 - A Resultados de gerações do prompt “Portrait of Woman”

Na segunda etapa, o mesmo conjunto de imagens foi utilizado no treinamento por meio de LoRA, desta vez na plataforma Scenario.ai e com base no Stable Diffusion XL. Diferentemente do DreamBooth, o LoRA operou como uma adaptação externa mais leve, que pôde ser combinada com modelos principais e outros treinamentos. Esse procedimento permitiu maior flexibilidade, especialmente na mistura entre estilo e identidade visual de personagens. Antes do treinamento, a plataforma forneceu descrições automáticas das imagens, geradas por visão computacional. Nesse momento, apareceram erros significativos de classificação: em algumas imagens, o berimbau foi identificado como esqui, vara de pescar ou espada. Embora essas descrições pudessem ser corrigidas manualmente, o problema evidenciou a dificuldade desses sistemas em reconhecer elementos específicos da cultura brasileira.



Figura 46 - Figura 4: Imagens geradas com o modelo utilizando o prompt “man with hat”

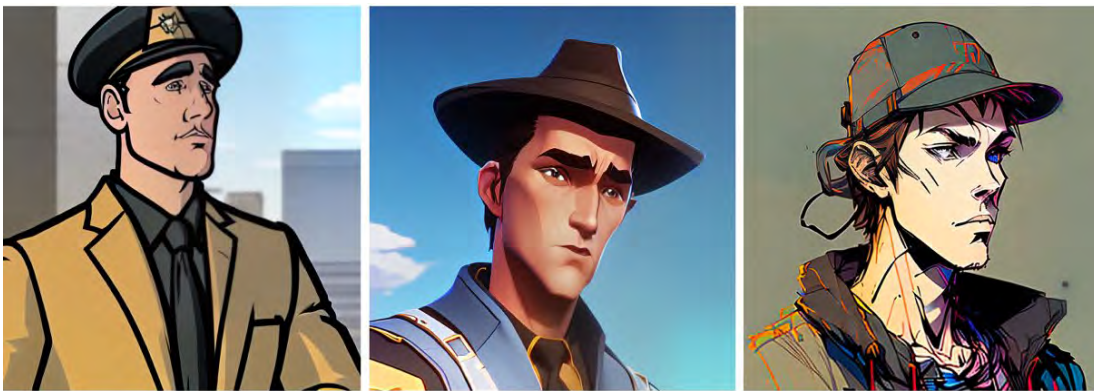


Figura 47 - Figura 35 - Figura 5: Imagens geradas nos modelos Archer Diffusion, Arcane Diffusion e SynthwavePunk v2 utilizando o prompt “man with hat”

Os resultados com LoRA mostraram maior aderência ao estilo original, tanto nos traços quanto em certas características faciais dos personagens representados. Além disso, a possibilidade de misturar esse modelo com outros treinados em fotografias de pessoas específicas abriu caminho para a geração de personagens relativamente consistentes em um mesmo estilo. Mesmo assim, permaneceram problemas de precisão cromática, pequenas variações de traço e diferenças de sombreamento, sugerindo que a consistência visual ainda dependeu de ajustes finos e controle cuidadoso dos pesos e combinações entre modelos.



Figura 48 - Resultados utilizando LORA e SDXL

A discussão derivada do experimento apontou que o estilo visual carregou consigo formas de representação. O que se aprendeu não foi apenas uma técnica gráfica isolada, mas também um modo de figurar corpos, rostos e contextos culturais. Em modelos genéricos, essa associação tendeu a reproduzir o apagamento de repertórios não hegemônicos, pois as bases de treinamento disponíveis em larga escala ainda foram

marcadas por predominâncias eurocêntricas e ocidentais. Quando essas imagens foram utilizadas para treinar novos sistemas, o efeito tendeu a se ampliar. O experimento mostrou, por contraste, que o treinamento com materiais localizados e culturalmente específicos pôde introduzir deslocamentos importantes, ainda que não resolvesse o problema de forma definitiva.

Essa observação também teve implicações para a discussão sobre autoria. O estudo não tratou as imagens geradas como equivalentes às ilustrações originais, mas como derivações produzidas a partir de um repertório visual previamente elaborado. O treinamento com material autoral apareceu, nesse contexto, como uma forma mais delimitada de uso da inteligência artificial, em comparação com modelos treinados sobre grandes acervos coletados sem critérios claros de consentimento ou contextualização. Mesmo quando as imagens geradas preservaram traços do estilo original, elas não reproduziram integralmente o olhar, a intenção e a sensibilidade envolvidos na produção manual. O estilo pôde ser parcialmente imitado, mas isso não significou que a experiência criativa do artista tenha sido transferida ao sistema.

Algum tempo após o experimento foi publicado um artigo a respeito de treinamento que deve ser mencionado, “Make it New”: AI and Modernism (Manovich; Arielli, 2023), que cita um experimento de Lev Pereulkov, com características semelhantes, trabalhando com construtivismo russo e pinturas. Hoje existem sistemas exatamente com esta função, o Exactly.ai oferece o serviço de treinamento de estilos para empresas gerarem infinitas ilustrações e a capacidade dos modelos foi aprimorada para isso.



Ao final deste experimento, dois anos após, a Adobe incorporou esta funcionalidade ao Firefly, seu modelo de produção de imagens generativas, e usamos o mesmo material para testar. Foi possível extrair retratos nos quais manteve o viés afrodescendente, mas na perspectiva etária o modelo procurou uma idade em torno de 30 anos, ao que parece. Foi possível ainda gerar algumas imagens de temas não treinados, soltos, aleatórios com estilo coerente ao treinamento em forma direta. Esses temas estarem bem representados demonstram a influência de um treinamento maior, de um modelo geral como o SDXL, na geração de imagens, em confluência ao treinamento de estilo que fizemos, operando de maneira bastante semelhante ao experimento com LORA. Ao tentar gerar um tocador de berimbau ou pandeiro, este modelo maior poderia ter esses assuntos treinados, mas pelo resultado, apesar de reconhecer, já que estava no treinamento, possivelmente não há dados suficientes no modelo grande.

Do ponto de vista do design, o experimento sugeriu que ferramentas generativas puderam ser incorporadas como recursos auxiliares na construção de linguagens visuais coesas, principalmente quando se desejou explorar variações dentro de um mesmo

universo gráfico. Isso pôde ser útil em contextos como ilustração seriada, desenvolvimento de personagens, sistemas visuais para jogos ou produção de imagens em sequência. Entretanto, a adoção dessas ferramentas exigiu avaliação crítica constante, tanto em relação aos vieses dos modelos quanto aos limites formais de seus resultados. No caso específico da série estudada, a complexidade dos movimentos corporais da capoeira, de seus instrumentos e de sua dimensão cultural, mostrou que a geração automática não substituiu satisfatoriamente o trabalho de construção visual desenvolvido nas ilustrações originais.

O espaço latente é o espaço de probabilidade, fruto de treinamento, sobre o qual é tokenizada a imagem em inteligência artificial generativa. Provê possibilidades praticamente infinitas, mas finito em variedade, visto que o que se apresenta parte de um universo de treinamento e diálogo com a máquina. É uma caixa obscura na qual podemos mexer, mas que conforme a abordagem pode ser contenção ou expansão da imaginação.

Por fim, o estudo permitiu pensar o treinamento de modelos generativos também como prática de letramento técnico e cultural. A experiência de treinar e gerar imagens não foi, em si, excessivamente complexa, e pôde ser utilizada em contextos educativos para discutir aprendizado de máquina, representação, acervo e cultura visual. Ao mesmo tempo, o avanço dessas ferramentas tendeu a ampliar a circulação de imagens sintéticas nas redes e nos ambientes digitais, o que pôde alterar os modos de produção, leitura e valorização das imagens em geral. Nesse cenário, tornou-se ainda mais importante manter abertos os processos, explicitar suas condições técnicas e políticas e sustentar a presença da avaliação humana diante da velocidade e da opacidade dos sistemas.

4.2 Preenchendo lacunas

Um exemplo de preenchimento dessas lacunas em processo manual foi o modelo do projeto V-Horus. Para essa experiência em VR, um modelo foi retocado, extraindo

texturas de fotografias de registro e aplicando-as sobre um modelo tridimensional em malha criado a partir de tomografia. Nesta situação, o processo foi manual, com recortes de detalhes das fotos aplicados sobre a malha, com o intuito de produzir um modelo realista que pudesse expressar a imagem do artefato perdido. Essa representação foi observada e avaliada por especialistas e conhecedores e obteve êxito. As ferramentas utilizadas foram de um fluxo de produção de modelos 3D, entre escultura digital para retoque, como Maxon Zbrush, texturização com Adobe Substance Painter e Blender para consolidar o modelo para implementação em Unity 3d. Uma vez em que os dados das texturas foram retirados de estudos de reconstrução do artefato, em detalhes, mas não com muitas vistas organizadas, o trabalho manual, com supervisão humana, criou consistência no processo.



Figura 49 - Modelo de tomografia texturizado e retocado manualmente para o projeto V-Horus

Nesses casos, modelos de IA generativa e visão computacional poderiam colaborar com reconstrução digital de artefatos. A informação visual esparsa, sendo organizada, poderia ajudar a preencher lacunas na construção da visualização tridimensio-

nal. Também é possível pensar sobre a quantidade de imagens necessárias para a criação de modelos para experiências e buscar recursos através para completar esses dados., podem ser ferramentas para a criação de experiências.

Uma diferença fundamental é que na utilização de IA os resultados são estimativas, no campo da probabilidade, observando uma provável complementariedade dentro do universo latente dos modelos. Essa característica se opõe ao modo determinístico tradicional das ferramentas de design em computação gráfica. A computação gráfica evoluiu a partir de coordenadas precisas e interfaces que possibilitam entradas de dados objetivos e entrega precisa dos resultados de operações, graficamente representados. De outro modo, neste modelo de geração probabilística, o resultado é fruto de um cálculo entre dados de entrada e treinamento do modelo. Dessa forma difere em essência, gerando um novo modo de trabalho, menos procedural por natureza.

Curiosamente, quando o modelo tenta gerar imagens com assuntos com os quais não foi treinado, pode “alucinar”. Como notaremos, a especificidade às vezes faz com que grandes modelos falhem. Por outro lado, a capacidade geracional é quase infinita e pode ser permeada por treinamento ou uso misto, aplicado dentro de processos de design, algo que já vem acontecendo na implementação de ferramentas de IA nos programas tradicionais de design como Adobe Photoshop.

4.3 Experimentos generativos 2D, treinamento e movimento

Este conjunto de observações de processos teve como objeto de três peças audiovisuais que combinaram dança e modelos generativos de difusão. Foram documentados no livro XR AI Realidades Expandidas e Inteligência Artificial (Azevedo et al., 2025). Eles partiram da observação de filmes em rotoscopia manual, como All Over de Paula Toller e Donavon Frankenreiter, realizado pela Comparsas Comunicação, assim como a observação da abertura da novela O Pai Ó (2) feita pelo Grupo Sal, ambos com pinturas de André Côrtes e Marcelo Gemmal. A questão que se apresentava era como se daria essa aproximação entre filme e geração de imagem por difusão e até que ponto esse processo poderia ser inserido como parte do fluxo de geração de filmes de animação com essa característica.

Em termos de método, os trabalhos variaram o “modo de guiar” oferecido ao modelo: no primeiro, a geração partiu de um vídeo como referência; no segundo, partiu de uma animação 3D derivada de captura de movimento; no terceiro, além de usar vídeo como guia, houve treinamento adicional para aproximar a saída de um estilo autoral. Em todos os casos, o resultado dependeu do arranjo do pipeline (preparo do vídeo, extração de frames, prompt, parâmetros e pós-produção), e não de um único ajuste isolado.

Do ponto de vista técnico, os experimentos operaram com Stable Diffusion e derivações, entendendo a difusão como síntese por etapas em que a imagem foi produzida pela remoção progressiva de ruído, guiada por texto e/ou imagem (Alammar, [S.d.]). Essa escolha implicou que o “controle” tenha permanecido distribuído entre a imagem de entrada (frames), o prompt e as tendências do próprio modelo.



Figura 50 - A imagem já com fundo recortado e sua interpretação através do Stable Diffusion, quadro por quadro

No experimento MOVE, a geração foi realizada a partir de um vídeo de dança, convertido em sequência de frames e processado via Deform sobre Stable Diffusion 1.5 (“deform-art/deform-stable-diffusion”, 2024). A direção visual buscou aproxi-

mações com repertórios modernistas associados a Kandinsky e com referências ao teatro e à dança da Bauhaus, incluindo o Balé Triádico (Schlemmer; Moholy-Nagy; Molnár, 2020). O desenvolvimento do material ocorreu por testes: primeiro em imagens estáticas (para estabilizar um vocabulário de formas e textura), e depois na sequência completa.



Figura 51 - Um sentido de agência na interpretação dos quadros

Foram aproveitadas as inconsistências entre frames, a difusão altera detalhes de um quadro para outro, visto que era produzido um por um. Esse “batimento” foi tratado

como parte do efeito visual do experimento, e não apenas como falha e acaba contrastando e sendo acentuado pelo movimento em câmera lenta. A finalização passou pelo Adobe After Effects (edição, sincronização e composição), com adição de texturas e ajustes gerais. A ênfase na diferença se relaciona com o efeito manual da roscopia, onde o trabalho de cada quadro em batimento ou é um defeito ou é parte da linguagem.

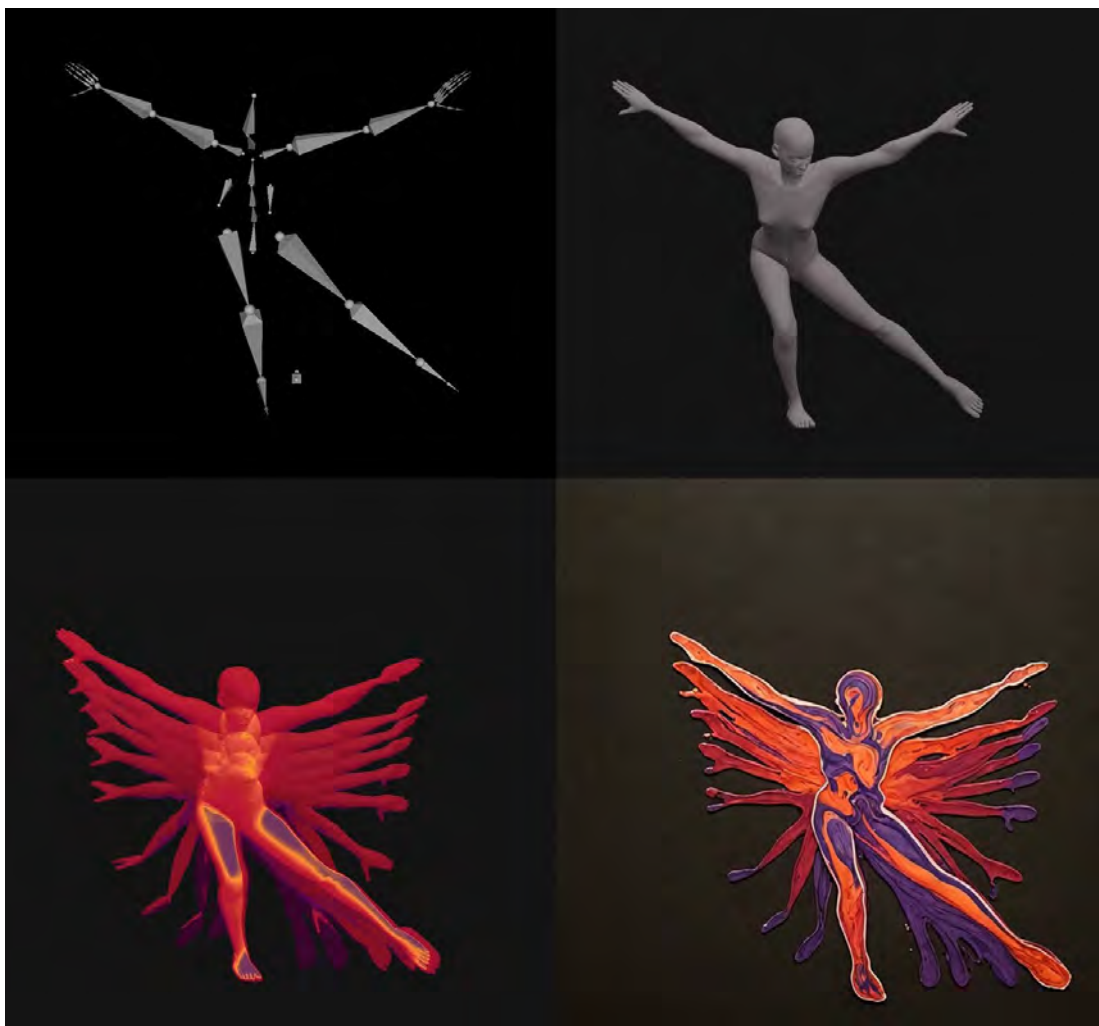


Figura 52 - Etapas da animação "CORES". Da esquerda para direita, a captura de movimento pura realizada no IMPA pelo grupo Space-XR. Um avatar criado no Blender através do complemento MBLab. O Terceiro um tratamento sobre o tempo realizado em after effects e o quarto a imagem gerada pelo Stable Diffusion.

No experimento CORES, o controle da forma foi deslocado para a construção prévia do movimento em 3D. A performance foi capturada por mocap, tratada no Blender e depois estilizada com SDXL em ferramenta de motion design. Um recurso central foi o uso do rastro temporal (“ecotemporal”), empregado para sugerir continuidade e energia do gesto. Esse uso do rastro dialogou com procedimentos de experimentação audiovisual em Pas de Deux (“Pas de Deux”, 1968).

Um prompt de render (“Colorful Suminagashi Marbling”) orientou a direção material, mas o experimento mostrou que a saída variou e produziu texturas que não corresponderam literalmente ao termo do prompt. Para reduzir instabilidades excessivas entre frames, foi testado o AnimateDiff (Guo et al., 2023), que buscou melhorar a consistência temporal sem exigir treinamento específico por cena. Como resultado, o experimento indicou que seria possível usar o sistema generativo como uma pós-produção de animações tridimensionais, viabilizando tratamentos de aparência orgânica.

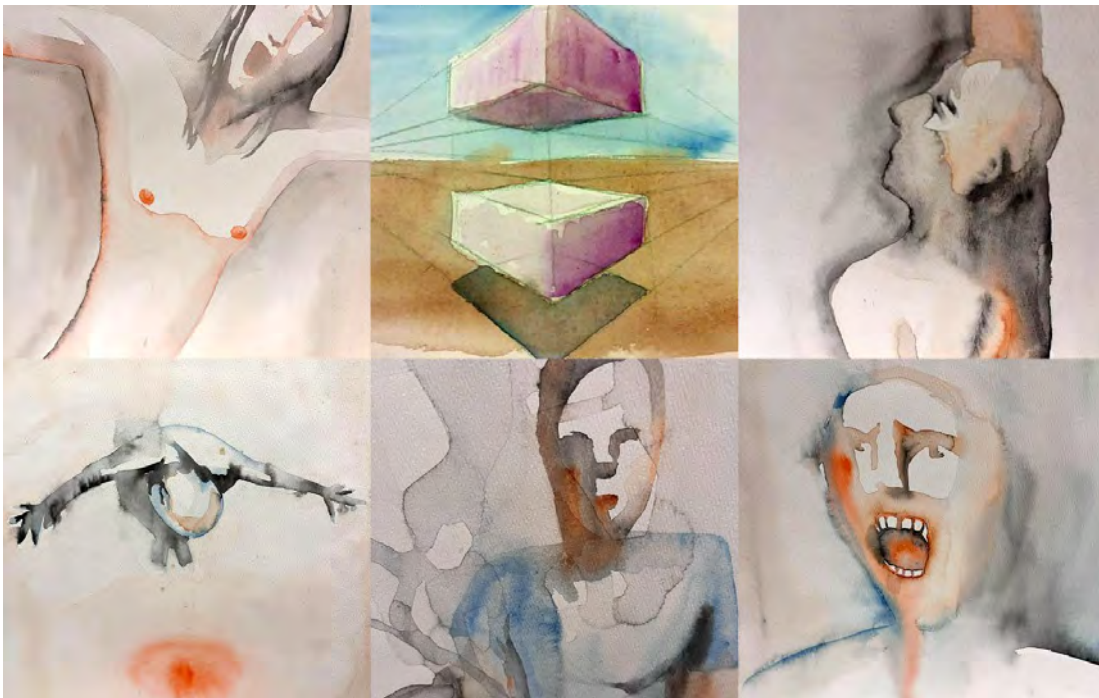


Figura 53 - Exemplo de imagens que fizeram parte do treinamento adicional utilizado em Gestos

No experimento GESTOS, além do vídeo-guia, buscou-se aproximar a saída de um estilo próprio por treinamento adicional. O pipeline utilizou SDXL com ComfyUI (“ComfyUI | Generate video, images, 3D, audio with AI”, [S.d.]), o que permitiu organizar a geração em nós e explicitar as etapas do processo. O ajuste fino foi feito via LoRA (Hu et al., 2021), a partir de um conjunto de imagens autorais.

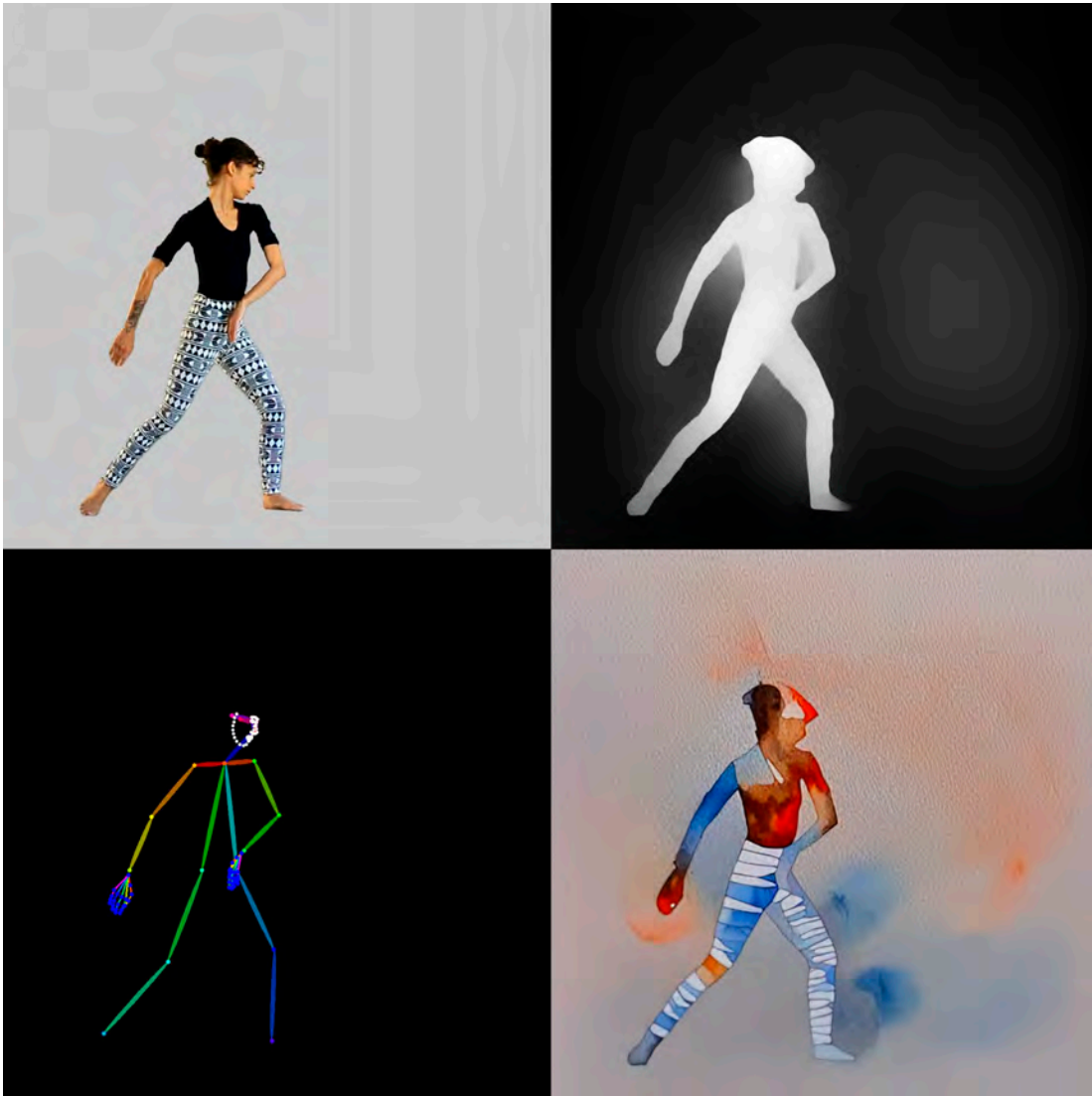


Figura 54 – A imagem original, o guia de controle de profundidade em cima, o guia de estrutura corporal embaixo e o resultado final. Os guias de controle para profundidade e estrutura corporal, MiDas e DWpose facilitam a reinterpretação da imagem.

O procedimento foi verificado em duas etapas: (1) geração por prompt para avaliar se o estilo tinha sido incorporado; e (2) geração guiada por frames para manter relação com o corpo em movimento. Para reduzir deformações e preservar estrutura corporal, foram introduzidos controles auxiliares: profundidade com MiDaS (Ranftl et al., 2019) e pose com DWpose (Yang et al., 2023), integrados por ControlNet (Zhang; Rao; Agrawala, 2023). O experimento também comparou alternativas de consistência temporal; quando a estabilização alterou demais a aparência pretendida, preservou-se parte do batimento entre quadros como característica do processo. A composição incluiu rotoscopia/recorte e ajustes de entrada, reconhecendo que condições simples (como fundo e contraste) alteraram o comportamento do modelo.

Os três experimentos têm em comum a exploração da relação entre material produzido com finalização generativa. Também faz parte do fluxo geral a geração de imagens estáticas antes da geração de múltiplos quadros. MOVE explorou a relação direta entre vídeo e difusão, incorporando variação quadro a quadro como parte da linguagem, trabalhando o aspecto generativo sobre a imagem como uma pós-produção manual. Ali foi possível avistar um sentido de agência artificial, como se fossem pintados milhares de quadros independentes por alguém, conforme uma solicitação coordenada por orientações. CORES deslocou o controle para o 3D, usando rastro e depois ferramentas para consistência em algo depois visto em uso do 3d como guia da geração de vídeo. GESTOS adicionou a dimensão de treino autoral e combinou isso com controles estruturais (profundidade e pose) para reduzir deformações sem “apagar” completamente as variações do modelo. O controle aparece como um equilíbrio entre modelo base, dados de referência e mecanismos auxiliares e os ajustes do pipeline fizeram parte do objeto analisado.

O experimento MOVE foi apresentado em exposição do FILE 2023, Festival Internacional de Imagem Eletrônica, no InShadow Festival em Lisboa, foi escolhido pelo Zinetika Festival para o prêmio Outras Luces.



Figura 55 - Imagens do filme criado inteiramente em IA

Após estes experimentos, onde o movimento era um recurso experimental em lógica de animação, os modelos de vídeo evoluíram em formato próprio. De uma maneira massiva foi iniciada uma corrida pelo vídeo em inteligência artificial. Como guias da geração, em geral seguiram o padrão de imagem e depois vídeo baseado em imagem, e posteriormente modelos de vídeo vêm implementando ferramentas de controle e algo que se parecem com *storyboard*, como Higginsfield e Google Flow.

A partir disso foram realizados alguns experimentos para verificar essa evolução enquanto vídeo e capacidade de representação, o que teve melhor consistência foi o VEO da Google. Foi observado que em pouco tempo não haveria diferença visível entre imagem generativa e imagem filmada, como já estava sendo estabelecido em imagens estáticas. Apesar da imagem fotográfica e o cinema já serem vistos como fruto de opiniões e processos, não um registro neutro, e as ferramentas digitais já terem capacidade de simulação anterior, a facilidade e fluidez do processo pareceu mudar o sentido da imagem digital. Na medida em que a simulação é facilitada, a força da imagem em movimento cai um pouco, algo parecido com a aura da obra de arte de Walter Benjamin, mas relativo à expectativa de índice que se esvai.

Parecia plausível obter algo semelhante ao conjunto de imagens para gerar 3DGS, e iniciamos alguns testes para gerar volume. Por outro lado, observamos também alguns modelos generativos tridimensionais, que no primeiro momento geravam quase ícones em 3D, mas que ao evoluírem poderiam ser uma forma de desenvolver

estudos ligados ao desenvolvimento de visualidade tridimensional. Porém antes cabia observar a possibilidade da geração tridimensional a partir de imagens, usando modelos feitos para este fim.

4.4 Objetos generativos 3D

Estes modelos são principalmente geradores de assets, isto é, objetos de composição de jogos e vídeos. No primeiro caso, experimentamos completar a tridimensionalidade utilizando inteligência artificial generativa, semelhante ao processo de difusão de imagens. Em geral, a forma de geração destes modelos é bastante ligada ao conceito de *assets* de jogos e funcionam principalmente gerando malhas. Meshy, Tripo, Trellis, e Huyen3d funcionam de maneira parecida em seu uso. Geralmente, são treinados com modelos e renderizações, por isso a aparência tem forte influência de estilos de modelagem tridimensional, mesmo com fotos como referência para o prompt., Por esse fator, é difícil sua utilização quando se trata de conteúdo específico. Possuem forte viés voltado para o que se espera de um modelo 3D. Apesar disso, porque aceitam mais dados de entrada, como várias imagens, conforme exemplo abaixo, podem ter resultados próximos.

Para observar o funcionamento, foi criado um experimento utilizando a escultura iconográfica Vênus de Willendorf. A escolha da peça foi pela simplicidade, pois parecia possível aos modelos e também pelas grandes quantidades de informações disponíveis para uso e para comparação. O plano foi usar as imagens de Bjørn Christian Tørrissen (“Venus of Willendorf”, 2025) ¹¹ das vistas da escultura pré histórica Venus of Willendorf (“Venus of Willendorf”, [S.d.]), descoberta por Josef Szombathy em 1908.

¹¹ Disponível em https://en.wikipedia.org/wiki/Venus_of_Willendorf#/media/File:Venus_of_Willendorf_-_All_sides.jpg acesso em 05/07/25



Figura 56 - fotos de Bjørn Christian Tørrissen

As vistas foram utilizadas como referência em todos os modelos para ter uma referência consistente do artefato. Os modelos foram Trellis 1 e 2, Huyuan 2 e 3, Meshy 6 e Hitem3D. Deve-se notar que a textura, característica importante desse modelo porque determina o material calcário oolítico, não foi representada apropriadamente.

Ainda que com ressalvas sobre a forma, a modelagem a partir da imagem é bastante convincente. No que diz respeito às texturas, a precisão é comprometida pela ausência de repertório e treinamento específico, ao que parece, seja pela representação do material em si, seja pela representação de inscrições, que são distorcidas e reinterpretadas. Seriam necessários modelagem manual usando as vistas como referência e ferramenta adicional como treinamento adicional específico ou a reaplicação manual da textura, usando Adobe Substance Painter, e as imagens de referência, como foi realizado no projeto V-Horus, por exemplo.



Figura 57 - Mostramos os modelos lado a lado em modo comparativo com: Um modelo retirado do Scan the City, Trellis 1, Trellis 2, na linha de baixo o Hitem3d e o Huyuan 3d 3 e o 2.

Porém, existe uma questão em relação à origem das imagens a partir da geração de inteligência artificial. Estes modelos e imagens são reconstruídos inteiramente. Desta maneira, o controle sobre a forma é perdido. Essa geração de artefatos confunde, pois perde indexação com a realidade, algo precioso ao processo de captura e divulgação. São inspirados em imagens, mas já são outras imagens. Ainda que retoque e transformações sejam feitos em outros modelos, existe índice do real em sua construção e uma cadeia mensurável de decisões, já neste caso é gerado um novo modelo a partir de

referência, o que em si não é problema, o problema é confundir uma versão com o original. Portanto, no caso de objetos em que se deseja preservar o índice, pode ser mais indicado proceder em torno do tópico anterior, extraindo profundidade e imagem, do que gerando algo novo.



Figura 58 - Figuras estilizadas inspiradas na Venus de Willendorf

Utilizar modelos deste método pode ser bastante generoso ainda para a criação de experiências estilizadas por sua característica generativa. Diferentes linguagens podem ser adotadas para ilustrar temas e para contar histórias, denotando diferentes tons para diferentes públicos. Essas ferramentas funcionam muito bem com entradas de pinturas digitais, especialmente se a imagem sugerir volume. Mesmo assim, devemos perceber os limites impostos pelo treinamento enviesado, isto é, estilos conhecidos, *low poly* e *cartoon* tridimensional como acima, funcionarão bem, estilos regionais, a exemplo de cerâmica brasileira, por exemplo, podem não funcionar, sendo dependentes de novos treinamentos (o Trellis da Microsoft, é *open source* e aceita treinamento).



Em um sentido conclusivo, para recuperação de visualidade a partir de fotos, existem modelos como o Reconviagen (Chang et al., 2025) que, conforme os modelos de visão descritos anteriormente, buscam reconstruir fidedignamente modelos 3D a partir de vistas, mesmo ainda não sendo exatos (aparece um efeito de moiré na textura, por exemplo) buscam consistência na representação, sem criar um novo modelo a partir da imagem de entrada. A intenção fundamental nesse caso é consistência com as imagens de entrada, uma característica de uma interpretação mais direta da imagem. Por isso, para muitas atividades de reconstrução modelos que não geram novas imagens, mas que extraem informações e as tratam, podem ter um resultado mais confiável, mesmo que incompleto.

4.5 Profundidade e significado / visão computacional

Aqui utilizamos recursos parecidos com os recursos de controle para gerar volumes a partir de imagens. O primeiro tipo a ser exemplificado são estimativas de profundidade geradas a partir de modelos pré-treinados que agem em torno da observação

do ponto de vista, detectando a posição da câmera e proximidade. São representados a partir de imagens, conhecidas como mapas de disparidade. Os mapas em si não são novos, existem como parte do processo como de aplicação de deslocamento em malhas em modelagem 3D. A maior parte dos modelos estudados foram experimentados em suas implementações para o Comfy UI, transportados de suas versões originais para o programa *opensource* por Andrea Pozetti.

A possibilidade de extrair de imagens diretamente é um recurso que vai desde tratamento de imagens no Adobe Photoshop até condução de carros autônomos. É possível ainda estimar o mapa de disparidade, a diferença entre a imagem vista em cada olho para produzir imagens estéreo e atribuir volume a imagens para levar ao tridimensional ou segmentar planos.

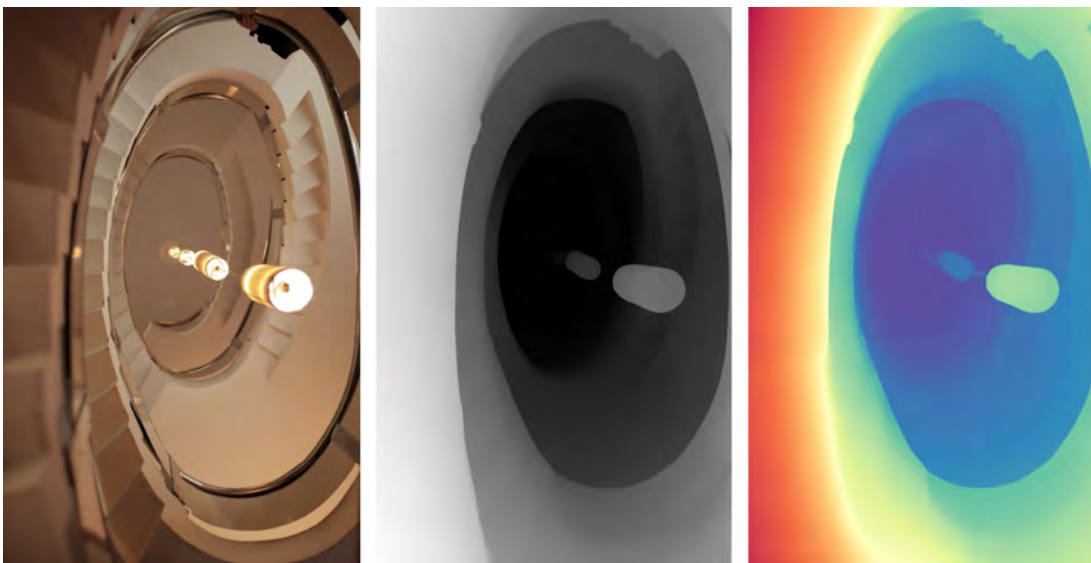


Figura 59 - Imagens geradas pelo modelo Depth Anything 2

Dessa forma percebemos ainda a possibilidade de uso gráfico, pois produz silhuetas separadas por cores. Dentro desta pesquisa foram feitas imagens de profundidade de modelos tridimensionais de temas abordados por diversos estudos do grupo XR-Space para a capa e projeto gráfico do livro XR-IA Realidades Expandidas e Inteligência Artificial e misturadas a malhas dos mesmos modelos, em parceria com o design

de Fabiana Prado, que utilizou as imagens como componentes do design. Estes modelos muitas vezes são incluídos em outros sistemas, como por exemplo para aprimorar o treinamento de 3DGS.



Figura 60 - Imagens de profundidade e sua aplicação gráfica

Quando a profundidade é extraída e mapeada com a imagem, começamos a visualizar uma nova possibilidade de tridimensionalidade. Realizamos testes com Depth Anything 3 e Apple Sharp, ambos geram 3DGS a partir de imagens bidimensionais.

O primeiro possibilita o uso de mais de uma vista e cruza as imagens para gerar uma nuvem de pontos (*pointcloud*) ou o 3DGS. Por triangular imagens buscando encaixar os pontos, linhas de perspectiva são bastante úteis. Para objetos soltos apresentou dificuldade em realizar a síntese, mas a proposta é bastante promissora em geral.

Já o Apple Sharp é realizado a partir de uma vista, monocular. Ele produz 3DGS em cores povoando densamente a partir dos pixels da imagem, ao que parece. Embora

a profundidade estimada não seja muito precisa, ele resolve a sensação de tridimensionalidade em um limite também chamado de 2.5D, que separa em planos e adiciona volume na direção da imagem, contemplando um ângulo de visão restrito. Para isso pode estimar a lente usando metadados da imagem ou por entrada do usuário. O modelo é extremamente rápido, leve e tem implementação simples. Como experimento foi utilizada diversas imagens como *Senhora sentada ao virginal* (Vermeer, 1670) de Johannes Vermeer e outras de Marc Ferrez disponibilizadas pelo Instituto Moreira Sales em licença *Creative Commons* para o Wikimedia.



Figura 61 - 3DGS criado a partir de pintura *Senhora sentada ao virginal* (1673) de Johannes Vermeer.

Muitos sistemas auxiliares ao treino de 3DGS têm sido produzidos, um modelo incorporado no treinamento de 3DGS dentro do Lichtfeld Studio é o ROMA (Edstedt et al., 2025), ele densifica adicionalmente a nuvem de pontos, estimando novos pontos a partir dos existentes, melhorando a qualidade do resultado e tornando o treino mais rápido, aumentando a precisão da representação.

Outros modelos de visão computacional podem reconhecer objetos e corpo humano. Estes modelos são capazes de saber o que está na imagem, segmentar, quantificar itens e pessoas, sempre dependendo de seu treinamento. O treinamento “ensina” o que deve ser observado, por isso quase inevitavelmente possui vieses. Reconhecer que há vieses é basal para a aplicação dessa tecnologia.

Ao caso de citar o pacote de modelos da Meta, SAM - *Segmentation Anything Model* (“SAM 3D”, [S.d.]) que identificam, segmentam e modelam objetos a partir de fotos. Embora a reconstrução de objetos as vezes não seja muito preciso com objetos que sejam levemente incomuns é eficaz para objetos cotidianos, reconhecidos pelo treinamento. Curiosamente, entrega arquivos em 3DGS com características de malha, sem brilhos ou transparências, entrega também arquivos de malha em GLB.

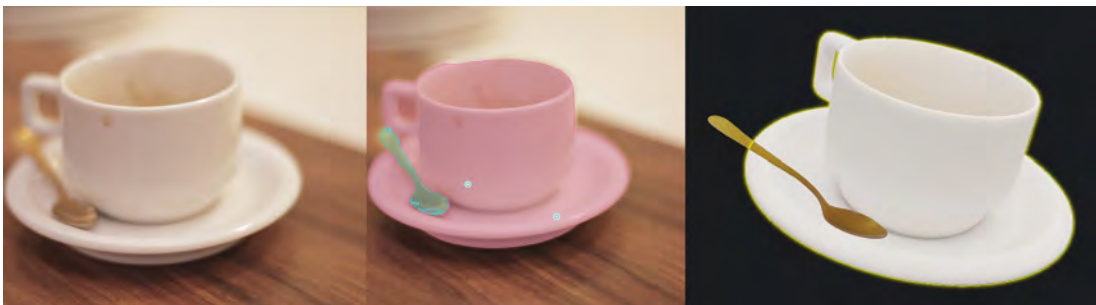


Figura 62 - O SAM detecta, segmenta e reconstrói objetos tridimensionais

Apesar de ser um modelo principalmente de visão computacional, considerando o nome segmentação, quando reconstrói a imagem está gerando uma nova imagem baseada em seu treinamento de significado. A geração da uma colher e da xícara, por exemplo é determinada pela identificação do objeto e sua construção, não usando a

referência da imagem e estimando volume, o que difere dos exemplos anteriores. Em sequência segmenta a imagem, identifica o objeto e suas características e estima a aparência, produzindo um modelo novo.

4.6 Vertendo o eixo temporal para profundidade

Aqui iniciam-se testes para geração de volume utilizando imagens bidimensionais de cunho histórico. Assim como toda mudança de suporte, não substitui a obra original, mas pode promover uma observação diferente para criar pontos de vista, nesse caso literalmente. Essa forma que foi testada difere das descritas acima por utilizar modelos de vídeo para o comando de rodar em torno do objeto, criar as vistas necessárias para a superfície por movimento e posteriormente treinar o 3DGS.

Nos primeiros testes existia muita dificuldade em obter coerência na rotação, principalmente quando envolvia pessoas, visto que os modelos de vídeo foram treinados com trechos de imagens em movimento, no entanto, esse movimento foi sendo facilitado conforme a evolução destes. O primeiro modelo em que foi possível executar foi a translação de um site comercial de vídeos generativos chamado Higginsfield em um modelo treinado com *bullet time*. Esse modelo imitava a câmera do filme Matrix (“Matrix”, 1999), provavelmente treinado a partir desse movimento de câmera e por isso aceitava rodar em torno de figuras humanas e objetos.



Figura 63 - 3DGS baseado em Natureza Morta com Flores. da Mota, Agostinho. Fonte: Acervo Pinacoteca de São Paulo, disponível em: <https://superspl.at/scene/2abdbccd>

A utilização de modelos de vídeo possibilitou obter uma rotação, de maneira bastante coerente, gerada a partir de grandes conjuntos de dados. A partir desta rotação foi possível criar imagens em 3DGS. São imagens fictícias, isto é, não existem no original, mas a imagem de referência entra como primeiro quadro, o que por si, já difere de modelos generativos tridimensionais. Essa coerência com a imagem de referência se deve aos conjuntos de dados gigantescos treinados em imagens reais ou mesmo diversas. O exemplo da pintura a óleo natureza-morta com flores de Agostinho da Mota (Da Mota, 1873), do acervo da Pinacoteca de São Paulo, é algo que funcionou de maneira singular. Trata-se de adicionar tridimensionalidade e possibilitar certa imersão nas imagens de modo que sejam observados detalhes e ampliar a percepção de realidade dessas obras.

Foram testados com fotografias e imagens geradas por modelos de IA generativa como Imagen da Google. Curiosamente, no primeiro momento, apesar das imagens geradas por IA generativa parecessem reais, falhavam após rotação enquanto as fotografias pareciam ter mais consistência geométrica.



Figura 64 - Tela verde e rotação utilizando Kling 3.1

Um processo semelhante foi realizado com o modelo WAN 2.2, open source, mas teve qualidade de imagem inferior, mostrando problemas comuns como rosto ou mãos deformados.

Com fotografias antigas, utilizamos a mesma ideia, mas sistematizando como uma captura intrínseca. Nesse tempo notamos que o modelo de vídeo obedecia a rotação, assim como o Grok, que tem como resultado uma resolução mais baixa. Testamos com o retrato de Machado de Assis, atribuído a Marc Ferrez, da coleção do Instituto Moreira Sales. Assim, a sequência segue recorte, aplicação de fundo verde (*chroma key*), para então entrar com o prompt: *“A video of a realistic statue , camera orbit 360 degrees around in perfect loop, stopped air, no bounce. green background Eyes are fixed. Body is fixed. Light is fixed. Face is fixed, Hair is fixed. Head is fixed. Clothes are fixed, hands and arms are fixed, It is a photographic immobile statue and camera orbit 360 degrees”* e com a imagem de referência fundo verde como início e fim do filme. O filme com a rotação é exportado e o fundo transformado em transparente em uma sequência de imagens .png. Então, a construção da nuvem de pontos é feita no Epic Reality que possui a qualidade de detectar coerência sem o fundo.

Após esta etapa, a nuvem é levada ao Lichtfeld Studio, densificada utilizando o Roma V2 e então transformada em Gaussian Splatting.



Figura 65 - Nuvem de pontos e câmeras criadas através do vídeo generativo e câmeras

Este processo cria um 3DGS usando o aspecto fotográfico da imagem, diretamente. Esse processo, baseado em vídeo, quando aceita a consistência do prompt em rodar tem um bom resultado, aparentemente porque tira proveito da larga base de treinamento dos modelos de vídeo. O ideal seria criar anéis por baixo e por cima apontando para o sujeito da imagem, em câmera inclinada, superior e inferior, mas isso ainda é um pouco complexo para coerência geométrica e para adesão ao prompt.



Figura 66 - Alguns ângulos do 3DGS, disponível em: <https://superspl.at/scene/4a00963b>

Uma ferramenta para melhorar a qualidade destas imagens é o ROMA (Edstedt et al., 2025) portado para o Lichtfeld Studio (Gmira, 2026), aplicado porque densifica com nuvens de pontos adicionais, o que embora não resolva a ausência de outras vistas, diminui os espaços vazios. Métodos semelhantes estão sendo testados, alguns mais voltados para a criação de modelos fictícios como outros com a camada generativa em comum.

Neste caso, estimamos volumes não existentes, mas capazes de criar interesse para exposições, cenários e outras aplicações acerca dos temas. Esse interesse pode ser justificável principalmente acompanhados de narrativas, como ilustrações tridimensionais.

5.

Representação humana e movimento

A representação do ser humano em ambientes digitais pode ser relacionada, de modo sintético, à história da arte ao se observar que cada regime de imagem reorganizou de maneira própria a figura humana, sua presença e sua relação com o observador. Na pintura, especialmente a partir da perspectiva renascentista, o corpo humano foi construído como centro da representação, articulando proporção, profundidade, gesto e expressão em um espaço visual organizado para o olhar. Nesse contexto, a figura humana não era apenas tema, mas também medida de inteligibilidade da imagem, associada a ideais de semelhança, presença e construção simbólica do sujeito (GOMBRICH, 1999).

Com a fotografia, essa relação se transformou, pois a imagem do corpo passou a ser vinculada a um registro técnico do real. A figura humana já não dependia apenas da interpretação manual do artista, mas de um aparato capaz de inscrever traços luminosos do mundo visível. Como observou Dubois (1993), a fotografia consolidou um tipo de imagem marcado pelo índice, isto é, por uma conexão física entre objeto e imagem. Ao mesmo tempo, Flusser (2002) mostra que essa aparente objetividade é sempre mediada pelo aparelho e por seus programas. Assim, a representação fotográfica do ser humano oscilou entre documento, memória, pose, encenação e construção cultural.

O cinema herdou da fotografia essa relação com o registro, mas acrescentou duração, movimento e narrativa. O corpo humano passou a ser representado não apenas como figura imóvel, mas como presença temporal, inscrita em ações, gestos e performances. Já nos ambientes digitais e na realidade virtual, esse processo avançou para outro estágio: o corpo pode ser escaneado, modelado, animado e experienciado em espaços computacionais, tornando-se menos um vestígio direto do real e mais uma

construção numérica e interativa. Desse modo, a representação do ser humano em ambientes digitais pode ser entendida como continuação e deslocamento da tradição da pintura, da fotografia e do cinema: da figura construída pela mão do artista, ao corpo inscrito pela luz, depois ao corpo em movimento e, por fim, ao corpo programável e imersivo (MANOVICH, 2001; GRAU, 2003). O corpo programável explica tanto o Instagram, cujos filtros alteravam a aparência, quanto avatares tridimensionais.

Humanos enxergam a si próprios com alto grau de exigência quando testam a realidade na representação. Nesta instância, observar que o conceito de vale da estranheza (Mori; MaDorman; Kageki, 2012). se aplica da mesma forma que com robôs, não só em relação a sua aparência, mas ao movimento e comportamento.

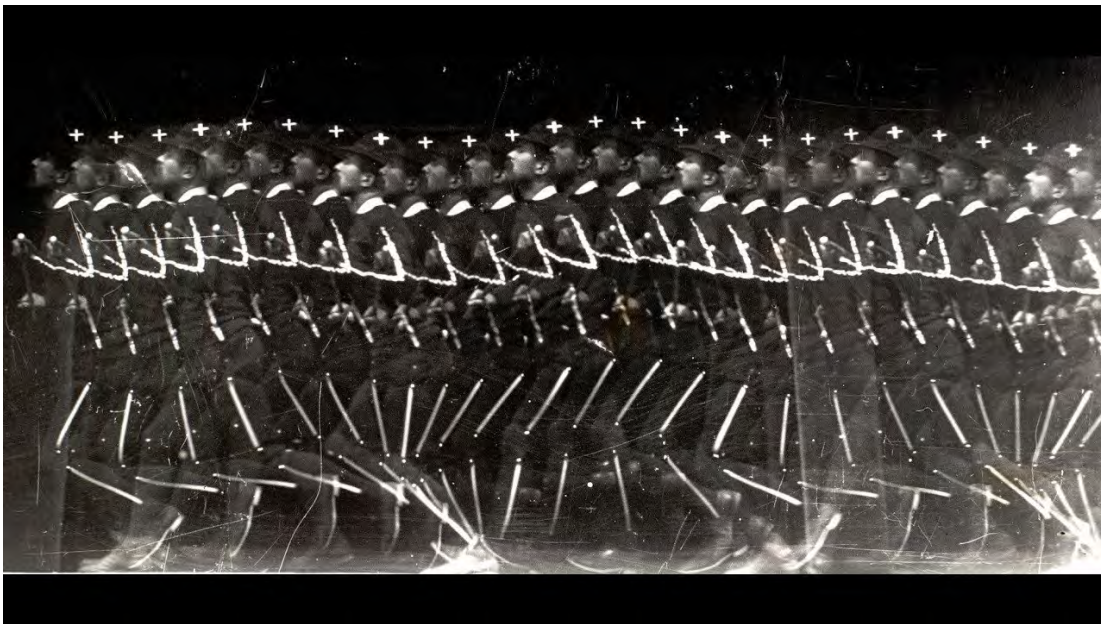


Figura 67 - Étienne-Jules Marey criou imagens cuja grafia inspirou a representação do mocap. Fonte: (Teston, 2004)

A captura de movimento (mocap) é uma técnica importante na passagem do corpo humano e gestos para o digital, a captura de movimento por sua vez tem a característica de separar os dados numéricos das coordenadas de movimentos do corpo. É essencial em fluxos de trabalho em animação em computação gráfica, utilizada em

muitas animações e efeitos em filmes. A captura de movimento remonta estudos de movimento de Eadweard Muybridge com suas cronofotografias e Étienne-Jules Marey que inventou graficamente algo muito parecido com o que conhecemos hoje como mocap (Kitagawa; Windsor, 2008).

O mocap segmenta o movimento da imagem em dados numéricos, funciona em geral para humanos, em alguns casos para animais. Torna o movimento quantificável, o que também torna essa informação útil em medicina e esportes.

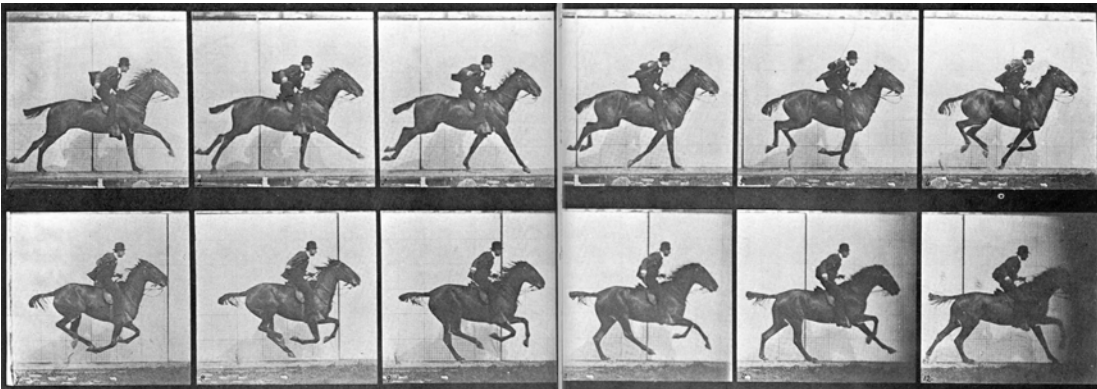


Figura 68 - Um dos primeiros estudos de movimento realizados por Muybridge para investigar o galope de cavalo. Fonte: (Q190568 et al., [S.d.])

Em ambientes imersivos temos avatares que usam a captura de movimento (mocap) em tempo real, em geral capturando imagens por câmeras do HMD e transferindo os movimentos para o avatar. A captura de movimento (mocap) contrasta de representações como a de 3DGS no quesito estrutural, enquanto o mocap busca coordenadas no espaço orientadas pelo movimento puro, e com isso esse movimento pode ser reaplicado sobre outros modelos, veremos que a direção do 3DGS em movimento, em geral, é o registro direto, ou como veremos, a aparência do registro direto.

Ao início de utilização dos metaversos em reuniões do grupo Space-XR durante a pandemia, para além de questões de aparência dos avatares, conforme examinaremos adiante, uma pergunta seria como o movimento humano poderia ser representado de forma realista. Naquele momento, os avatares tinham movimentos bastante limitados

e incorretos, apesar das câmeras dos HMDs serem capazes de mapear os pés, por sua localização. Somente a parte superior do corpo e o rosto funcionavam. Apesar de usar um sistema de avatares realista, a captura de movimento muitas vezes distorcia a posição dos braços, por exemplo.

Um primeiro esboço de movimento de corpo inteiro para visualização foi realizado utilizando movimentos importados do aplicativo/website Mixamo (“Mixamo”, [S.d.]), inteiramente feito no Blender 3D e exportado para o sistema Spatial.io em uma plataforma simplificada.

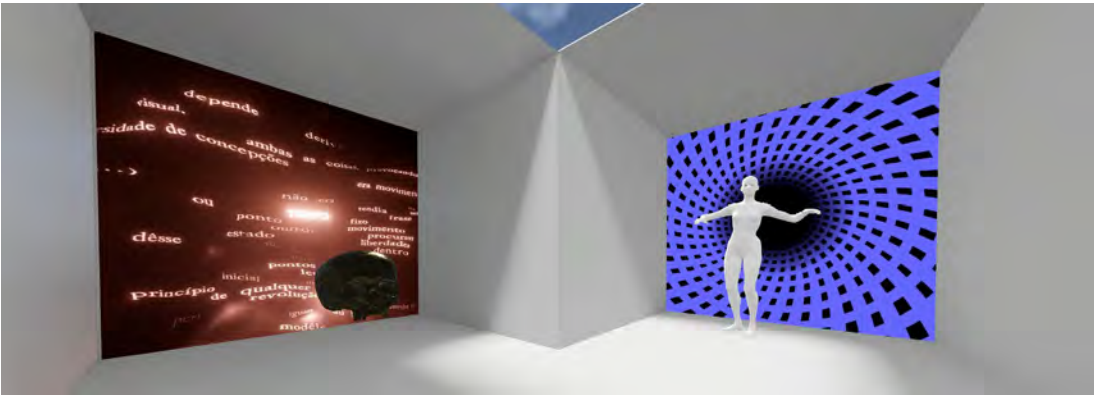


Figura 69 - Espaço de experimentos em metaverso com modelo tridimensional com movimento importado do mixamo para o Blender 3d.¹²

5.1 Ocean Dance

Conforme a descrição no artigo *Ocean Dance* grupo Space-XR iniciou pesquisas em torno de dança no VISGRAF Lab no IMPA, através de performances da pesquisadora Thaisa Martins e captura utilizando criamos alguns protótipos. Após alguns testes, foi feito um metaverso no Spatial.io que acabou dando o nome para o grupo de experiências.

¹² Disponível em <https://www.spatial.io/s/cube-629911c0ff451100014ed8f5>

Para a criação da instalação audiovisual pelo grupo Space-XR “Ocean Dance” foram criados alguns metaversos utilizando o Unity 3D e a plataforma Spatial.io, a partir da performance de Helena Bevilaqua e Fabio Nunes, com movimentos capturados no laboratório VISGRAF Lab no IMPA. A captura foi realizada utilizando sensores Optitrack e gerado arquivos BVH, com as coordenadas do corpo em movimento. Esses arquivos de movimento foram importados no Blender 3D e retocados e transferidos para um modelo criado por um *addon* de humanóides, o Mb lab. Foi então transferido para o Unity com acesso ao SDK do Spatial.io.



Figura 70 - Registro da captura de movimento de Thaisa Martins (com Luiz Velho em diálogo sobre a captura em curso), em sequência, Fabiano Nunes e Helena Bevilaqua.

O trabalho foi apresentado no Seminário Danc3 & T3tch, no Rio de Janeiro e em Porto Alegre. Algumas observações puderam ser feitas a partir do deslocamento dos participantes no espaço virtual, poucos participantes de fato rodaram em torno dos modelos dançantes; sobre a interação com o espaço, as rampas foram a melhor forma de mudança de altura. Algo a se reparar sobre o ponto de vista dos metaversos, diferentemente de observar um modelo 3d, onde podemos rodá-lo, para ver seu topo por exemplo, se tudo estiver no mesmo plano a altura também será constante.



Figura 71 - Imagem da experiência com performance de Helena Bevilaqua capturada no VIS-GRAF/IMPA¹³.

Assim, o observador de um metaverso ou de um ambiente que simule presença tem a altura quase fixa, pode abaixar, mas dificilmente deita-se no chão, dificilmente avista acima do modelo, se este for da mesma altura ou mais alto. Essa mudança de ponto de vista no ambiente virtual pode ser provocada por mudanças no ambiente.

Capturar, avaliar o material capturado, retocar e endereçar os esqueletos observando uma captura de precisão ajudou a criar critérios para avaliar outras formas de captura. A partir desse processo, também foram elaborados estudos gráficos em wireframe que acabaram inspirando o material de registro em dança enviado para o livro.

¹³ Disponível em: <https://www.spatial.io/s/xr-dance-experience-Helena-Bevilaqua-64447b5a696c55107064c555> e <https://www.spatial.io/s/xr-dance-experience-Fabiano-Nunes-644a96304882e3688772df3b>

5.2 Mocap IA

A captura inteligente, sem marcadores, envolve visão computacional para localizar. O corpo humano padrão tem formas identificáveis e através de visão computacional. Câmeras comuns ou equipamentos como o Lidar ou Microsoft Kinect, podem fornecer dados de posição, bidimensionais, no caso de câmeras comuns ou tridimensionais por equipamentos que percebem profundidade. Em nosso caso, a câmera é mais interessante por ser um objeto comum nos dias de hoje.



Figura 72 - Exploração gráfica através de captura de movimento

Alguns desses sistemas, como o Mediapipe (Lugaresi et al., 2019), funcionam em tempo real, mas em geral os sistemas que entregam dados com qualidade para animação necessitam de um tempo de processamento. A preferência por software livre nos fez observar o projeto Freemocap (Project, [S.d.]), que como *open source* facilitou a compreensão dos processos em funcionamento. Por exemplo, sabemos que o projeto começou triangulando posições em multicâmeras obtidas através do Mediapipe e depois trocou por Yolo, um outro modelo de visão computacional. Foram realizadas vários testes e usos. Entre os modelos testados estão Meshcapade, DeepMotion, Plask (Figura 64), Move Ai, Rokoko Vision, Mimem Ai e QuickMagic.

Em um experimento vídeo concebido em parceria com Matteo Moriconi realizamos uma demonstração no Instagram do VFXRio, com vídeo fornecido por ele, em um momento de surgimento da tecnologia, realizado em Wonder Dynamics. Esse sistema específico tem a peculiaridade de capturar o movimento, remover automaticamente o fundo, inclusive com a câmera em movimento. A maior parte dos sistemas de mocap através de vídeo indica a câmera parada como uma das premissas de boa captura.

No entanto, os melhores resultados foram obtidos no Quickmagic (“Quick Magic AI Mocap - Any Video, Any Motion”, [S.d.]) e no Move Ai em versão de uma câmera e posteriormente Mimem.ai. Enquanto o primeiro usa o vídeo enviado como referência e calcula outras posições, além de mixar relações físicas do movimento com o chão, o segundo possibilita a entrada de várias câmeras, resolvendo questões de oclusão de partes do corpo durante o movimento. Os vídeos são gravados por câmeras separadas e sincronizados no aplicativo. Nesse caso, gravado com três câmeras; podem ser utilizadas até doze, o que certamente garante maior precisão.

O movimento capturado é um rabo de arraia, também conhecido como meia lua de compasso, movimento original de capoeira. Na medida em que é um movimento que inverte pés e mãos passa a ser um bom teste para a capacidade de leitura de posição.



Figura 73 - Movimento capturado por 3 cameras utilizando mimem.ai

Outro modelo que requer atenção é o SAM3D Body, embora ainda funcione por poses estáticas, é um modelo capaz de perceber dados corporais de múltiplas pessoas em uma mesma imagem. No exemplo é usado a gravura Cabocle Indien civilisé (Debret, 1834). A vantagem específica de detectar vários pessoas na cena é que expande a utilização para avaliação e captura de dados através de imagens cotidianas. Existe inclusive a possibilidade de expansão para a utilização da segmentação em movimento (Gao; Miao; Han, 2025).



Figura 74 - Utilização do modelo SAM 3D Body para avaliar imagem

Para representar a possibilidade de mocap generativo, podemos citar aqui o sistema Kimodo (Rempe et al., 2026) da NVIDIA, criado para geração de movimentos. Extremamente eficiente para movimentos comuns, conectando os movimentos em uma interface detalhada.

Como modelo generativo treinado, segue o treinamento ainda básico, e responde erroneamente a movimentos não treinados, como movimentos de capoeira ou dança contemporânea. Foi realizado um breve experimento no qual foram utilizados prompts com nome de movimentos como ginga, tão comum que é contemplado no Mixamo, ou Laban square. Os resultados foram distantes dos prompts apresentando

kick boxing e algo semelhante a *hip hop*, onde podemos ler a tentativa de aproximação em classe de movimento, luta e dança, mas a ausência de referência específica.

Porém o sistema viabiliza interpolação de posições personalizadas, o que viabiliza a criação de movimentos não treinados. O que caracteriza uma interpolação inteligente que pode ser muito útil para animadores e mesmo para captura em tempo real.

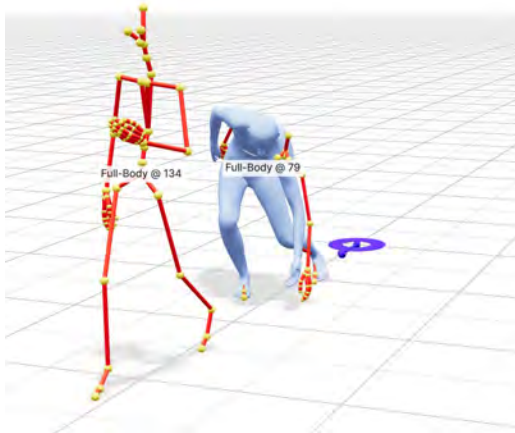


Figura 75 - O sistema Kimodo funciona com prompts e posicionamento dos rigs

5.3 Avatares

A representação da presença humana em espaços virtuais por meio de avatares inicia-se historicamente com o desenvolvimento dos ambientes digitais interativos, especialmente desde os MUDs e mundos virtuais das décadas de 1980 e 1990, nos quais o usuário passou a atuar no espaço informacional por meio de uma figura textual ou gráfica que o substituíva simbolicamente. Nesse contexto, o avatar não funcionou apenas como imagem de identificação, mas como forma de corporificação mediada, isto é, como interface pela qual o sujeito se apresentava, agia e estabelecia relações com outros em ambientes digitais. Como observou Turkle (Turkle, 1996), esses espaços favoreceram experimentações de identidade e de projeção do eu, enquanto Taylor (2002) destacou que os avatares se tornaram centrais para a constituição da experiência social e incorporada nos mundos virtuais. Em outras palavras, a presença humana em tais

ambientes deixou de depender do corpo físico compartilhado e passou a ser construída por sinais visuais, comportamentais e simbólicos codificados digitalmente.

Do ponto de vista conceitual, a presença por avatares pode ser entendida como uma forma de incorporação digital, em que o usuário experiencia o espaço virtual a partir de um corpo substituto que organiza percepção, ação e interação. Essa discussão também se articula à noção de presença social, entendida como a sensação de estar com o outro em um ambiente mediado (Biocca; Harms; Burgoon, 2003). Em ambientes de realidade virtual, realidade aumentada e realidade misturada, situados no continuum realidade-virtualidade proposto por Milgram e Kishino (1994), o avatar assume diferentes graus de mediação, podendo variar de uma representação simples até corpos virtuais complexos e responsivos. Assim, o avatar pode ser compreendido simultaneamente como dispositivo técnico, forma visual e construção cultural: ele representa o humano no espaço virtual, mas também redefine o que significa estar presente, agir e ser percebido em contextos digitais.

Convivemos hoje com representações adaptadas ao meio, com massivas informações videográficas e fotográficas, muitas vezes filtradas em redes sociais assim como figuras estilizadas entre emoticons e personas virtuais representando empresas. A representação digital é cotidiana, mas não tridimensional.

Em ambientes tridimensionais, a pesquisa participou empiricamente de um momento de adoção e uma oscilação entre a busca de uma representação realista para uma estilizada modular nas plataformas de realidade virtual.

A plataforma virtual com maior base de usuários é o VR Chat. Nesta plataforma, a liberdade de criação de avatares é total, isso fez, em anos de disponibilidade e uma abordagem plural, que existisse todo tipo de representação. Algumas experiências com essa plataforma foram feitas, captura de movimento ao vivo, no VISGRAF, a construção de um espaço semelhante ao *Ocean Dance* e a construção de um avatar. Em geral é semelhante ao Spatial.io, utilizando o Unity 3D e o SDK distribuído pela plataforma. Uma característica dessa plataforma é a diversidade de representações, com personagens de anime, antropro zoomorfismo, por vezes desproporcionais e com

todo tipo de expressão gráfica. Sendo uma plataforma pioneira, viabilizou a personalização de avatares através de kit de desenvolvimento (SDK), hoje com muitas ferramentas e exemplos no Blender, para modelagem e no Unity 3D, motor de jogo onde a plataforma é desenvolvida.

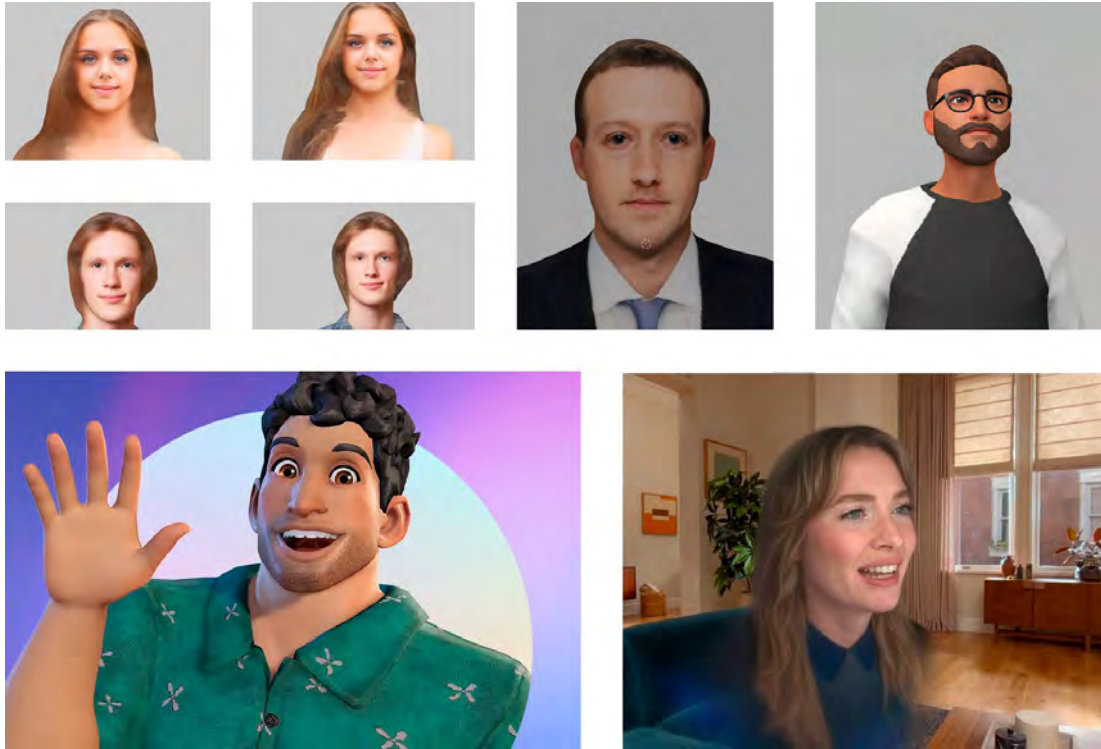


Figura 76 - Representação por sistemas de avatar, na linha de cima, cinco exemplos do Avatar SDK e do Ready Player Me e na de baixo, avatar da Meta e da Apple, usando 3DGS.

Quando o grupo Space XR começou a usar o sistema Spatia.io, a plataforma utilizava o sistema AvatarSDK que buscava mapear a imagem fotográfica sobre o modelo 3d. Este foi substituído aos poucos pelo Ready Player Me, que indicava uma representação que se desviava do realismo. Talvez esse movimento tenha feito parte de uma mudança de direção do Spatial.io, que no início se apresentava como espaço para reuniões virtuais, exposições e aulas e depois direcionou-se bastante a jogos online. O Ready Player Me era personalizável e poderia ter itens vendáveis.

Ready Player Me foi plataforma de avatares 3D interoperável que permite criar um personagem personalizado a partir de uma selfie para uso em mais de 8.000 aplicativos e jogos. Durante aproximadamente quatro anos esteve presente em muitas plataformas de jogos e realidade virtual, incluindo o Spatial.io. A plataforma oferecia modelos leves e infraestrutura para implementação, incluindo uma ampla possibilidade de personalização. Baseada em malha, texturas leves e rigging, possui uma lógica modular de construção, entre as partes do corpo, e um aspecto não realista conhecido, semelhante a personagens adultos da Pixar. Os avatares eram compatíveis com Unity, Unreal Engine e diversas plataformas, facilitando a consistência da identidade digital em mundos virtuais, VR, e experiências sociais. Em 2025 a plataforma foi comprada pela Netflix, de modo que está parando de operar em todos os sistemas anteriores.

Porém, conforme um estudo *Do realistic avatars make virtual reality better? Examining human-like avatars for VR social interactions* (Fraser et al., 2024) avatares realistas conferem maior credibilidade durante experiências sociais imersivas. De fato, o que vemos em redes sociais em duas dimensões não são muitas pessoas utilizando avatares não realistas. Com exceção de alguns momentos em que há ondas comportamentais ligados a algum acontecimento, como a geração de imagens do Studio Ghibli através de IA generativa, em geral as pessoas usam suas próprias fotos.

Avatares parecem mudar de paradigma, a Apple em seu Apple Vision Pro implementou uma solução chamada Personas em 3DGS (“Radiance Fields”, [S.d.]) juntamente com inteligência artificial, em um sistema que fotografa o rosto em vários ângulos e cria um avatar articulado realista. Nesta aplicação, os avatares voltam a ser uma redução do corpo, apenas cabeça e mãos. Não se sabe exatamente a construção que foi feita, mas desde 2023 a Apple tem pesquisas sobre o uso de 3DGS como, por exemplo, HUGS: Human Gaussian Splats (Kocabas et al., 2023), mas nesse caso eram avatares de corpo inteiro inicializados por SPML. Então a solução parece bastante com um rosto com rigging facial, como GaussiansAvatar (Qian et al., 2023). Isso importa porque mostra o quão mistas essas soluções podem ser.

Apesar de não possuir índice em tempo real, pois o avatar é uma síntese em IA realista, a aparência do real, junto com sincronia de gestos e voz, tenta preencher este

espaço no ambiente tridimensional. A ideia de avatares em gaussian splatting também é apresentada em outros sistemas para o Apple Vision Pro para agentes de IA em uma espécie de NPC em TaoAvatar: Real-Time Lifelike Full-Body Talking Avatars for Augmented Reality via 3D Gaussian Splatting (Chen et al., 2025), este sistema propõe agentes em IA para o HMD e para dispositivos móveis com apresentação realista e pode colaborar em narrativas de modo semelhante ao feito por Luiz Velho, Sergio Azevedo e Jorge Lopes durante apresentação de V-Horus 2.0 na Semna – Semana de Egip-tologia do Museu Nacional 2021 (Museu Nacional UFRJ, 2021), ao apresentarem um avatar em homenagem póstuma ao professor Dr. Antonio Brancaglione Junior.

5.4 Pessoas e movimento

A representação de pessoas é uma das principais qualidades do 3DGS, com índice real. Para cenas estáticas a solução costuma ser usar vídeo na captura para operar mais rapidamente e muita luz para evitar desfoque de movimento.



Figura 77 - Imagem gerada em 3DGS, disponível em:

Mas a solução industrial são sistemas de captura volumétrica. Domos geodésicos com dezenas ou centenas de câmeras sincronizadas, luzes e *chroma key* podem capturar várias posições simultaneamente. Esses sistemas já existem anteriormente ao formato 3DGS, para captura de volumes em malha, em movimento. Metastage, por exemplo, possui plugins para implementar em Unity 3D, Unreal e etc. Uma referência de vídeo volumétrico disponível no Oculus Quest é o teatro musical imersivo “All Kinds of Limbo”(“All Kinds of Limbo”, 2021), com Nubya Brandom como intérprete das músicas cantando em tempo real. Esse tipo de captura volumétrica em 3DGS chama-se 4DGS, por acrescentar a dimensão temporal. O Gracia (“Gracia”, [S.d.]), aplicativo para Metaquest, que começou criando visualização para 3DGS, tem implementado soluções de compressão e streaming em 3D.



Figura 78 - Um vídeo em IA foi utilizado como teste para gerar um 4DGS utilizando o Apple Sharp

Uma outra opção para a visualização com índice e em movimento seria a captura da imagem e criar volume, como faz o Apple Sharp ML. A partir da disponibilidade desse sistema no ComfyUI, foi realizada uma pequena alteração, onde ao invés da interpretação volumétrica de uma imagem, poderia ser feita toda uma sequência de

quadros, gerando um vídeo volumétrico 2.5D. Nesse caso a imagem só possui a dimensão frontal preenchida, mas tem profundidade estimada. O modelo é veloz na conversão o que poderia, em pouco tempo, se tornar uma visão tridimensional ao vivo. A ideia foi adotada pelo programador Andrii Shramko para o desenvolvimento de um pequeno software independente, distribuído *open source* que teve larga adoção. Durante a própria pesquisa, buscando formas de fazer transmitir o 4DGS encontrei a sugestão de uso deste próprio programa.



Figura 79 - O 4DGS gerado através de frame de vídeo e IA

Com a percepção de avanço de compreensão de prompt e controle de câmera nos modelos de vídeo mais recente, como demonstrado no item 4.6, chegamos a testar a possibilidade de gerar 4DGS a partir de sequência de quadros, que rodados, podem treinar as gaussianas e serem colocadas em sequência. Esse treino ainda é bastante rústico, mas preservando o ponto de vista central e em movimento é possível usar uma

aproximação ao pixilation, variação com elementos humanos do stop-motion consagrada por Norman McLaren em *Neighborhoods*. Desse modo, algo que parece difícil e caro, como a captura tridimensional de pessoas, pode ser acessível. Considerando a velocidade do desenvolvimento da geração de vídeo através de IA, esse fluxo pode ser simplificado.

6.

Considerações Finais

Ao longo deste trabalho, o 3D Gaussian Splatting foi observado como uma forma de representação tridimensional que contribui para repensar a construção da imagem em experiências imersivas. Em vez de depender principalmente de modelagem por malhas, UVs e texturas, esse procedimento organiza a imagem a partir da aparência, da luz, da cor e da variação do ponto de vista. Em alguns sentidos, toda tecnologia envolvida é dependente das imagens captadas e oferecidas como referência e material de treinamento. Com isso, aproxima-se de questões já presentes na fotografia e em outras imagens técnicas, sobretudo quando o problema envolve presença, observação e medição. De certo modo, existe um antagonismo em capturar a realidade em torno e utilizar IA para a produção de imagens. A captura de 3DGS se parece até mesmo com a fotografia analógica, pois no momento da tomada de imagens ainda não existe previsão do resultado, revelado no treinamento. Nessa perspectiva, seu interesse não esteve apenas em aspectos de desempenho, mas também em como esse formato interfere nas escolhas projetuais ligadas à construção de objetos, ambientes e corpos em XR.

Ao mesmo tempo, o percurso desenvolvido indicou que o 3DGS não deve ser entendido como substituto geral de outras formas de modelagem. A representação por malhas continua pertinente em situações que exigem maior controle estrutural, simplificação geométrica, animação ou compatibilidade com fluxos já consolidados. Já o 3DGS tende a responder melhor em casos nos quais a aparência capturada, as irregularidades de superfície, os reflexos, a vegetação e certos efeitos de presença visual têm maior importância. Assim, a questão não é estabelecer uma hierarquia fixa entre técnicas, mas reconhecer que diferentes problemas de projeto pedem diferentes regimes de imagem. Esse ponto recoloca a escolha técnica no campo do design, entendido como seleção de meios em função de objetivos específicos.

Esse aspecto também ajuda a explicitar a interdisciplinaridade presente no trabalho. Ao longo da pesquisa, design, computação gráfica, visão computacional, inteligência artificial, patrimônio, audiovisual e práticas artísticas apareceram articulados de forma contínua. Essa articulação não funcionou como soma ocasional de áreas, mas como condição do próprio processo de investigação. O grupo Space-XR, reuniu pesquisadores de diferentes centros e formações e desenvolveu experimentos em realidade virtual, realidade aumentada, captura de movimento, medicina, divulgação científica e expressão artística. Nesse contexto, o design operou como instância de articulação entre métodos, linguagens, ferramentas e objetivos de comunicação, contribuindo para organizar um campo que, por natureza, já se apresenta como híbrido.

No caso da inteligência artificial generativa, o trabalho mostrou que seu uso precisa ser tratado com atenção. Ela amplia possibilidades de variação, preenchimento, reinterpretação e síntese de imagens, vídeos e objetos, mas também pode alterar a relação entre imagem e evidência. Quando a geração probabilística passa a fazer parte do processo de construção visual, torna-se necessário distinguir com mais clareza o que deriva de captura, o que resulta de reconstrução e o que pertence ao campo da inferência. Momentos de verificação, como *storyboard*, *layouts*, consultas e participação de outras pessoas são importantes em uma perspectiva projetual. A potência do recurso aparece na capacidade de geração de interpolações possíveis, que se desdobram em técnicas diversas. Para criar movimentos, imagens ou principalmente para criar uma terceira dimensão em imagens ou pequenos conjuntos de imagem.

É possível pensar em uma consistência necessária entre a solicitação da geração, o treinamento realizado e o resultado. Caso essa consistência falhe, ou teremos alucinações do modelo de IA ou teremos uma mediocrização do resultado. Para compreender a geração de imagens é preciso entender as características do treinamento. Isso pode ser feito investigando através de documentação disponível, ou mesmo observando os resultados obtidos. Pode inclusive dispensar conhecimento técnico em IA mas é necessário apuro no olhar e no propósito. O uso sistemas de geração de modelos 3D, visão computacional e de vídeo acaba por elucidar essa diferença. Enquanto o treinamento de sistemas de modelos 3d é feito na comparação entre render e modelagem, sistemas

de visão computacional tentam encaixar índices para extrair tridimensionalidade e os modelos de vídeo são treinados em imagens reais, então se parecem com a matéria prima necessária para a criação do 3DGS.

Esse ponto é especialmente importante em contextos de interesse histórico-cultural, nos quais a imagem pode ser recebida como documento. Por isso, os experimentos com inteligência artificial foram tratados como parte de um conjunto de operações que exigem mediação crítica e explicitação de intenção.

Nesse contexto, a questão da aura do objeto pode ser compreendida em tensão com seu índice. A captura tridimensional, assim como a fotografia, preserva um vínculo ótico com o referente, uma vez que parte dos dados são obtidos a partir do encontro entre câmera, luz e objeto. Nesse sentido, permanece um componente de índice, isto é, uma relação com algo que esteve efetivamente diante do dispositivo. No entanto, essa condição não equivale à transferência integral da aura do objeto, entendida aqui como sua singularidade, sua presença situada e sua inserção em um contexto histórico e material específico. Ao ser convertido em imagem navegável, modelo digital ou experiência imersiva, o objeto pode preservar traços de sua existência, mas passa a operar em outro regime de presença, marcado pela reprodutibilidade, pela circulação ampliada e pela mediação técnica. Porém quanto mais a digitalização mantém aspectos de índice do real mais recompõe a experiência de aura do original. Dessa forma muda e influencia a experiência que pode ser provocada.

A visão computacional aparece no trabalho como recurso técnico importante para segmentar, estimar profundidade, reconhecer poses, classificar elementos visuais e converter imagens em dados operáveis. No entanto, esses processos não são neutros. Ao definir o que conta como corpo, objeto, fundo, movimento ou profundidade, a visão computacional também produz recortes e simplificações. Isso se torna mais evidente quando a passagem do bidimensional ao tridimensional depende de descrições automáticas, mapas de profundidade e reconhecimento corporal. Dessa forma, essas ferramentas ajudam a construir a imagem ao mesmo tempo em que delimitam o que pode ser reconhecido, descrito e processado. Por outro lado, ampliam possibilidades de des-

dobramentos, quantificação e estimativa. Nos momentos onde há informação suficiente, é capaz de estimar movimento, volume e interpolar inteligentemente informações visuais.

Essa discussão ganha clareza nos experimentos com representação humana. O trabalho indicou que a presença do corpo em XR não depende apenas de sua aparência visual, mas também de gesto, deslocamento, ritmo e temporalidade. A captura de movimento, nesse sentido, não aparece apenas como recurso complementar, mas como modo de tratar dimensões do humano. Isso ficou evidente nos experimentos com performance e dança, nos quais o problema da representação não se limitou à forma tridimensional, mas envolveu também comportamento, ação e relação espacial. A imagem do humano, portanto, exige articular aparência e movimento, o que amplia a discussão para além da superfície visível.

A presença da capoeira e da dança desloca essa discussão para o campo do patrimônio imaterial. Nos experimentos descritos, essas práticas não funcionam apenas como tema visual, mas como conjuntos culturais em que corpo, gesto, música, instrumento, memória e transmissão coletiva estão ligados. Por isso, ajudam a evidenciar um limite recorrente dos processos generativos e classificatórios: certos elementos podem até ser parcialmente reproduzidos como aparência, sem que sua densidade cultural seja efetivamente representada. O caso do berimbau, identificado em alguns momentos como esquí, vara de pescar ou espada, sintetiza esse problema ao expor uma dificuldade de reconhecimento de objetos específicos da cultura brasileira. Nesse sentido, capoeira e dança mostram que o patrimônio imaterial exige formas de representação que considerem também contexto, prática, temporalidade e sentido cultural, e não apenas forma visível.

Por fim, esse conjunto de questões permite retomar o design no sentido de desígnio. Ao longo do trabalho, o design não apareceu apenas como campo da forma final, mas como espaço em que se organizam critérios, intenções e limites de cada processo. É nesse nível que se decide quando usar captura direta, quando recorrer à reconstrução, quando aceitar inferência generativa e quando explicitar a diferença entre documento e interpretação. É também nesse nível que se articulam patrimônio material e imaterial,

objeto e performance, imagem e movimento, além da tensão entre índice e aura na digitalização de objetos culturais. Assim, a relevância do 3DGS, da IA generativa, da visão computacional e da captura de movimento depende menos de cada ferramenta isoladamente e mais da maneira como elas são integradas em uma prática projetual capaz de lidar com mediação, contexto cultural e responsabilidade na representação.

É possível que o formato 3DGS se transforme, mas é necessário reconhecer que sua viabilidade e distribuição aberta criaram um cenário de desenvolvimento em termos de estrutura da representação tridimensional digital. Apesar do surgimento ter sido a partir da captura, talvez como formato de representação, incluindo o sintético, venha a modificar todo um cenário já estabelecido.

Nesse sentido a nova representação em 3DGS é bastante generosa ao aproximar de um registro fotográfico. Uma comparação entre vetor e bitmap, em uma perspectiva do design, é útil para compreender as diferenças entre os formatos em malha ou campo de radiâncias, ainda que não suficiente para caracterizar uma forma final de visualização. Esse formato pode mudar, a escala de resolução, por exemplo, sempre muda conforme potência de processamento gráfico, por isso, tanto sintético quanto real convém conservar arquivos originais.

A conjunção com inteligência artificial pode fazer com este formato torne viável a integração entre imaginar e capturar. Para captura do ser humano, pode tornar imagens bidimensionais em tridimensionais, facilitando a criação de vistas, interpolando movimentos e viabilizando o 4DGS, oferecendo uma nova forma alternativa à captura em domos de câmeras. Além disso novas formas de manipulação da imagem surgem, como selecionar por assunto ou modificar luminância.

Nessa perspectiva, a inteligência artificial redistribui a rede de decisões envolvida na produção projetual. Diante de processos parcialmente automatizados, passam a fazer parte do projeto a escolha de bases de dados, a definição de objetivos, a elaboração de instruções, a seleção de saídas, a curadoria de resultados e a avaliação de efeitos. Por isso, a noção de design enquanto desígnio torna-se ainda mais relevante, já que o ato projetual deixa de consistir apenas na configuração direta de formas e passa

a envolver também a definição de intenções, critérios, parâmetros, mediações e contextos de uso. A inteligência artificial, nesse sentido, pode ser entendida como mais uma ferramenta para a produção de experiências, imagens, programas e processos, ainda que introduza novos níveis de opacidade técnica e decisória.

Nesse contexto, o viés não aparece apenas como problema externo ao sistema, mas como parte das condições de seu funcionamento. Ele pode ser tensionado, deslocado ou parcialmente corrigido quando confrontado com imagens de referência, treinamentos adicionais e escolhas mais específicas de base visual. No entanto, isso exige lidar com a opacidade dos sistemas de modo crítico, com clareza de propósito e leitura consciente dos resultados produzidos. Mais do que buscar neutralidade plena, trata-se de compreender que toda geração automatizada envolve recortes, exclusões e ênfases, o que reforça a necessidade de mediação projetual.

Uma possibilidade de enfrentar esses desafios culturais está no diálogo com outras mídias e com materiais oriundos do mundo real. Imagens de referência, conjuntos de treinamento adicionais, modelagem tridimensional, registros fotográficos e outras formas de intervenção podem funcionar como modos de orientar ou tensionar os resultados gerados. Nesse processo, torna-se visível uma questão mais ampla: a convivência entre o corpo programável e o corpo que programa. O envio de imagens, a troca de dados e o treinamento específico mostram que o corpo não é apenas objeto de representação nessas ferramentas, mas também um agente que orienta, ajusta e condiciona os sistemas. Também é preciso considerar que os suportes técnicos continuam em transformação. HMDs e outros dispositivos atuais ainda podem ser entendidos como protótipos funcionais de visualização tridimensional. Eles já permitem experiências novas, em alguns casos ainda pouco familiares ao público, e servem a treinamento, simulação e fruição imersiva. No entanto, sua popularização em larga escala ainda depende de caminhos técnicos, econômicos e culturais que permanecem em aberto.

7.

Referências Bibliográficas

180 years of 3D | Royal Society. Disponível em: <<https://royalsociety.org/blog/2018/08/180-years-of-3d/>>. Acesso em: 30 jan. 2026.

360 Extractor tool for Blender. Disponível em: <<https://toppinappi.gumroad.com/l/360extractor>>. Acesso em: 18 mar. 2026.

ALAMMAR, Jay. **The Illustrated Stable Diffusion.** Disponível em: <<https://jalammarmar.github.io/illustrated-stable-diffusion/>>. Acesso em: 5 fev. 2025.

All Kinds of Limbo. Disponível em: <<https://www.allkindsoflimbo.com>>. Acesso em: 31 mar. 2026.

AMORY, Dita. **Georges Seurat (1859–1891) and Neo-Impressionism - The Metropolitan Museum of Art.** Disponível em: <<https://www.metmuseum.org/essays/georges-seurat-1859-1891-and-neo-impressionism>>. Acesso em: 9 fev. 2026.

ARRIVAL.SPACE. **Arrival.Space.** Disponível em: <<https://arrival.space/>>. Acesso em: 28 mar. 2026.

AZEVEDO, Sérgio *et al.* **Realidade Expandida: Novas mídias e IA.** Tradução: Jorge Lopes. Rio de Janeiro, RJ: Editora Mourthé, 2025.

BASZUCKI, David. The Metaverse is coming. **Wired**, 2 jan. 2021.

Bilateral Guided Radiance Field Processing. Disponível em: <<https://bilarfpro.github.io>>. Acesso em: 10 fev. 2026.

BIOCCA, Frank; HARMS, Chad; BURGOON, Judee K. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. **Presence: Teleoperators and Virtual Environments**, v. 12, n. 5, p. 456–480, out. 2003.

BOLTER, Jay David; GRUSIN, Richard. **Remediation: understanding new media.** 5.ª ed. Cambridge, Mass.: MIT Press, 2002.

BUCHANAN, Richard. Wicked problems in design thinking. v. 8, n. 2, p. 5–21, 1 abr. 1992.

Caixão de Sha-amun-en-su - Egito Antigo | Museu Nacional - UFRJ. Disponível em: <<https://museunacional.ufrj.br/dir/exposicoes/arqueologia/egito-antigo/arqegit009.html>>. Acesso em: 7 mar. 2026.

CHANG, Jiahao *et al.* **ReconViaGen: Towards Accurate Multi-view 3D Object Reconstruction via Generation.** arXiv, , 2025. Disponível em: <<https://arxiv.org/abs/2510.23306>>. Acesso em: 11 mar. 2026

CHEN, Jianchuan *et al.* **TaoAvatar: Real-Time Lifelike Full-Body Talking Avatars for Augmented Reality via 3D Gaussian Splatting.** arXiv, , 2025. Disponível em: <<https://arxiv.org/abs/2503.17032>>. Acesso em: 30 mar. 2026

ComfyUI | Generate video, images, 3D, audio with AI. Disponível em: <<https://www.comfy.org/>>. Acesso em: 8 mar. 2026.

CSIKSZENTMIHALYI, Mihaly. Flow: The Psychology of Optimal Experience. *In: [S.l.: S.n.]*.

DA MOTA, Agostinho. **Natureza-morta com flores.** , 1873. Disponível em: <<https://acervo.pinacoteca.org.br/online/ficha.aspx?id=9641&ns=201000&lang=br&c=pesquisa&IPR=5254>>

Dany Bittel, Macro Splats. Disponível em: <<https://danybittel.ch/macro>>. Acesso em: 2 fev. 2026.

DEBEVEC, Paul E.; MALIK, Jitendra. Recovering high dynamic range radiance maps from photographs. *In: THE 24TH ANNUAL CONFERENCE. Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97.* Not Known: ACM Press, 1997. Disponível em: <<http://portal.acm.org/citation.cfm?doid=258734.258884>>. Acesso em: 26 mar. 2026

DEBRET, Jean Baptiste. **Cabocle Indien civilisé.** , 1834. Disponível em: <<https://acervo.pinacoteca.org.br/online/ficha.aspx?id=9472&ns=201000&lang=BR&c=pesquisa&IPR=1602>>. Acesso em: 30 jan. 2026

deforum-art/deforum-stable-diffusion. deforum-art, , 18 jan. 2024. Disponível em: <<https://github.com/deforum-art/deforum-stable-diffusion>>. Acesso em: 18 jan. 2024

DEUTSCH, Isaac *et al.* **PPISP: Physically-Plausible Compensation and Control of Photometric Variations in Radiance Field Reconstruction.** arXiv, , 2026. Disponível em: <<https://arxiv.org/abs/2601.18336>>. Acesso em: 7 mar. 2026

Edmund Fraser Studio. Disponível em: <<https://www.edmundfraser.co/>>. Acesso em: 2 fev. 2026.

EDSTEDT, Johan *et al.* **RoMa v2: Harder Better Faster Denser Feature Matching.** arXiv, , 2025. Disponível em: <<https://arxiv.org/abs/2511.15706>>. Acesso em: 7 mar. 2026

ENGINE. **KIRI Engine: 3D Scanner App for iPhone, Android, and Web - KIRI Engine.** Disponível em: <<https://www.kiriengine.app/>>. Acesso em: 13 fev. 2026.

Eric Paré - Montréal - **bullet-time / light-painting / gaussian splat photographer.** Disponível em: <<https://ericpare.com/>>. Acesso em: 26 jan. 2026.

Experimenting with Light Fields. Disponível em: <<https://blog.google/products-and-platforms/products/google-ar-vr/experimenting-light-fields/>>. Acesso em: 18 mar. 2026.

FLUSSER, Vilém. **Filosofia da caixa preta : ensaios para um futura filosofia da fotografia.** [S.l.]: Annablume, 2011.

FRASER, Alan D. *et al.* Do realistic avatars make virtual reality better? Examining human-like avatars for VR social interactions. **Computers in Human Behavior: Artificial Humans**, v. 2, n. 2, p. 100082, ago. 2024.

GAO, Mingqi; MIAO, Yunqi; HAN, Jungong. **SAM-Body4D: Training-Free 4D Human Body Mesh Recovery from Videos.** arXiv, , 2025. Disponível em: <<https://arxiv.org/abs/2512.08406>>. Acesso em: 2 abr. 2026

GMIRA, Shady. **shadygm/Lichtfeld-Densification-Plugin.** , 27 mar. 2026. Disponível em: <<https://github.com/shadygm/Lichtfeld-Densification-Plugin>>. Acesso em: 30 mar. 2026

GODIN, Danny; ZAHEDI, Mithra. Aspects of Research through Design: A Literature Review. **DRS Biennial Conference Series**, 16 jun. 2014.

GORTLER, Steven J. *et al.* The lumigraph. *In: SIGGRAPH96: 23RD INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques.* ACM, ago. 1996. Disponível em: <<https://dl.acm.org/doi/10.1145/237170.237200>>. Acesso em: 23 jan. 2026

Gracia. Disponível em: <<https://www.gracia.ai/>>. Acesso em: 31 mar. 2026.

GRAU, Oliver. **Virtual Art: From Illusion to Immersion.** Cambridge, MA, USA: MIT Press, 2004.

GUO, Yuwei *et al.* **AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning.** arXiv, , 2023. Disponível em: <<https://arxiv.org/abs/2307.04725>>. Acesso em: 22 mar. 2026

HU, Edward J. *et al.* **LoRA: Low-Rank Adaptation of Large Language Models.** arXiv, , 16 out. 2021. Disponível em: <<http://arxiv.org/abs/2106.09685>>. Acesso em: 5 fev. 2025

Jawset Postshot. Disponível em: <<https://www.jawset.com/>>. Acesso em: 13 fev. 2026.

KANDINSKY, Wassily; BRANDÃO, Eduardo; KANDINSKY, Wassily. **Ponto e linha sobre plano: contribuição à análise dos elementos da pintura.** São Paulo: Martins Fontes, 2005.

KAZHDAN, Michael; HOPPE, Hugues. Screened poisson surface reconstruction. **ACM Transactions on Graphics**, v. 32, n. 3, p. 1–13, jun. 2013.

KERBL, Bernhard *et al.* **3D Gaussian Splatting for Real-Time Radiance Field Rendering.** arXiv, , 8 ago. 2023. Disponível em: <<http://arxiv.org/abs/2308.04079>>. Acesso em: 22 jan. 2024

KITAGAWA, Midori; WINDSOR, Brian. **MoCap for artists: workflow and techniques for motion capture.** Amsterdam ; Boston: Elsevier/Focal Press, 2008.

KOBIE, Nicole. The Dream of the Metaverse Is Dying. Manufacturing Is Keeping It Alive. **Wired**, 5 maio 2025.

KOCABAS, Muhammed *et al.* **HUGS: Human Gaussian Splats.** arXiv, , 29 nov. 2023. Disponível em: <<http://arxiv.org/abs/2311.17910>>. Acesso em: 23 jan. 2024

La Motilla del Azuer - Daimiel, Spain - Download Free 3D model by Global Digital Heritage and GDH-Afrika (@GlobalDigitalHeritage). , 4 set. 2018. Disponível em: <<https://sketchfab.com/models/750c1dc950274be39705545f4a492bd9/embed?autostart=1>>. Acesso em: 21 mar. 2026

LARIC, Oliver. **Three D Scans. Selfportrait (Photosculpture)**, [S.d.]. Disponível em: <<https://threedscans.com/albertina/francois-willeme-1/>>. Acesso em: 10 dez. 2025

LEVOY, Marc; HANRAHAN, Pat. Light field rendering. *In*: SIGGRAPH96: 23RD INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES. **Proceedings of the 23rd annual conference on Computer graphics and interactive techniques.** ACM, ago. 1996. Disponível em: <<https://dl.acm.org/doi/10.1145/237170.237199>>. Acesso em: 23 jan. 2024

LI, Xiaoqiao *et al.* Reimagining Creativity: Gaussian Splatting as an Organological Extension of Imagination in Artistic Processes. **Leonardo**, v. 58, n. 3, p. 241–249, 1 jun. 2025.

LOEW, Katharina. Tangible Specters: 3-D Cinema in the 1910s. v. 37/38, n. 31, p. 87–116, 2013.

LUGARESI, Camillo *et al.* **MediaPipe: A Framework for Building Perception Pipelines.** arXiv, , 2019. Disponível em: <<https://arxiv.org/abs/1906.08172>>. Acesso em: 27 mar. 2026

Luma AI - Interactive Scenes. Disponível em: <<https://lumalabs.ai/interactive-scenes>>. Acesso em: 13 fev. 2026.

MANOVICH, Lev; ARIELLI, Emanuele. ‘Make it New’: GenAI, Modernism, and Database Art. 2023.

Matrix. Nashville, Tennessee, EUA Warner Bros., , 1999.

Metropolitan Museum of Art. Acervo. Disponível em: <<https://www.metmuseum.org/pt/art/collection/search?showOnly=has3d&offset=0>>. Acesso em: 1 abr. 2026.

MICROSOFT. **The Next Rembrandt: Recreating the work of a master with AI.** **Microsoft News Centre Europe**, 13 abr. 2016. Disponível em: <<https://news.microsoft.com/europe/features/next-rembrandt/>>. Acesso em: 23 mar. 2026

MILDENHALL, Ben *et al.* **NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.** arXiv, , 2020. Disponível em: <<https://arxiv.org/abs/2003.08934>>. Acesso em: 23 jan. 2026

MILGRAM, Paul *et al.* Augmented reality: A class of displays on the reality-virtuality continuum. **Telemanipulator and Telepresence Technologies**, v. 2351, 1 jan. 1994.

MILGRAM, Paul; KISHINO, Fumio. A Taxonomy of Mixed Reality Visual Displays. **IEICE Transactions on Information Systems**, v. Vol E77-D, p. 1321–1329, 12 dez. 1994.

MITCHELL, Vinicius; NOVAES, Luiza. Experiências e ambientes interativos. *In: Formas do Design: 30 Anos do Programa de Pós-graduação em Design PUC-Rio.* 1. ed. Rio de Janeiro, Brasil: Rio Books, 2024. p. 201–218.

Mixamo. Disponível em: <<https://www.mixamo.com/#/>>. Acesso em: 24 mar. 2026.

MORI, Masahiro; MACDORMAN, Karl; KAGEKI, Norri. The Uncanny Valley [From the Field]. **IEEE Robotics & Automation Magazine**, v. 19, n. 2, p. 98–100, jun. 2012.

MUSEU NACIONAL UFRJ. [VIII SEMNA] **Conferência 2: V- Horus: a persistência da memória através das imagens.** , 1 dez. 2021. Disponível em: <<https://www.youtube.com/watch?v=rAk0j4AY6LQ>>. Acesso em: 2 abr. 2026

PARK, Sunyoung. The Conceptual Study of Photosculpture. **Culture & History Digital Journal**, v. 14, n. 1, p. 289, 11 jul. 2025.

Pas de Deux. National Film Board of Canada, , 1968. Disponível em: <https://www.nfb.ca/film/pas_de_deux_en/>. Acesso em: 12 jun. 2025

PETT, Daniel. **BritishMuseumDH/irthorruCartonnage: A release of data to recreate mummy cartonnage in 3D.** Zenodo, , 25 jan. 2018. Disponível em: <<https://zenodo.org/record/1160620>>. Acesso em: 9 fev. 2026

PROJECT, The FreeMoCap. **FreeMoCap - Free Motion Capture for Everyone.** Disponível em: <<https://freemocap.org>>. Acesso em: 28 mar. 2026.

Q190568 *et al.* **Fichier:Muybridge horse gallop.jpg — Wikipédia.** Disponível em: <https://commons.wikimedia.org/wiki/File:Muybridge_horse_gallop.jpg>. Acesso em: 2 abr. 2026.

QIAN, Shenhan *et al.* **GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians.** arXiv, , 2023. Disponível em: <<https://arxiv.org/abs/2312.02069>>. Acesso em: 30 mar. 2026

Quick Magic AI Mocap - Any Video, Any Motion. Disponível em: <<https://quick-magic.ai/home>>. Acesso em: 2 abr. 2026.

Radiance Fields. Disponível em: <<https://radiancefields.com/apple-confirms-personas-use-gaussian-splatting>>. Acesso em: 30 mar. 2026.

RANFTL, René *et al.* **Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer.** arXiv, , 2019. Disponível em: <<https://arxiv.org/abs/1907.01341>>. Acesso em: 22 mar. 2026

REALIDADE EXPANDIDA. Disponível em: <<https://visgraflabimpa.br/xr-ai/portugues/>>. Acesso em: 22 jan. 2026.

REIS, Luciano. **Internet 3D com VRML.** Disponível em: <<https://web.tecgraf.puc-rio.br/~reis/vrml/>>. Acesso em: 13 mar. 2026.

REMPE, Davis *et al.* **Kimodo: Scaling Controllable Human Motion Generation.** arXiv, , 16 mar. 2026. Disponível em: <<http://arxiv.org/abs/2603.15546>>. Acesso em: 2 abr. 2026

ROMBACH, Robin *et al.* **High-Resolution Image Synthesis with Latent Diffusion Models.** arXiv, , 13 abr. 2022. Disponível em: <<http://arxiv.org/abs/2112.10752>>. Acesso em: 18 jan. 2024

SAM 3D. Disponível em: <<https://ai.meta.com/research/sam3d/>>. Acesso em: 14 mar. 2026.

SCHLEMMER, Oskar; MOHOLY-NAGY, László; MOLNÁR, Farkas. **The theater of the Bauhaus.** Tradução: Arthur S. Wensinger. Zurich: Lars Müller Publishers, 2020.

SCHÖN, Donald A. **The reflective practitioner: how professionals think in action.** New York: Basic Books, 1983.

SCHONBERGER, Johannes L.; FRAHM, Jan-Michael. Structure-from-Motion Revisited. *In*: 2016 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR). **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Las Vegas, NV, USA: IEEE, jun. 2016. Disponível em: <<http://ieeexplore.ieee.org/document/7780814/>>. Acesso em: 4 fev. 2026

Sha-Amun-en-su. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Sha-Amun-en-su&oldid=71465309#cite_note-Revista_Fapesp_Sha_Amun_en_su-2>. Acesso em: 25 mar. 2026.

SIGGRAPH (CONFERENCE) (22ND : 1995 : LOS ANGELES, Calif). *et al.* **Computer graphics : proceedings : SIGGRAPH 95 Conference proceedings, August 6-11, 1995.** [S.l.]: New York, NY : ACM SIGGRAPH, 1995.

SIMON BETHKE. **Tutorial - Gaussian Splatting.** , 21 out. 2024. Disponível em: <<https://www.youtube.com/watch?v=08NYHDwOqow>>. Acesso em: 18 mar. 2026

simonbethke. Disponível em: <<https://superspl.at/user?id=simonbethke>>. Acesso em: 7 fev. 2026.

Spatial - Create Immersive UGC, Virtual Classrooms, Experiential Marketing. Disponível em: <<https://www.spatial.io/s/Tartaruguito-62ceb78ca5a20c0001a43bd4?share=3193191752720264133>>. Acesso em: 23 mar. 2026.

STEPHENSON, Neal. **Snow Crash.** [S.l.]: Editora Aleph, 2021.

Stereotypography: typical, even racist, typefaces. Disponível em: <<https://grapeine.com/en/magazine/stereotypography-typefaces-racist/>>. Acesso em: 21 mar. 2026.

SuperSplat - The Home for 3D Gaussian Splatting. Disponível em: <<https://superspl.at>>. Acesso em: 4 fev. 2026.

TAYLOR, T. L. Living Digitally: Embodiment in Virtual Worlds. *In*: SCHROEDER, Ralph (Org.). **The Social Life of Avatars.** Computer Supported Cooperative Work. London: Springer London, 2002. p. 40–62.

TESTON, Bernard. L'œuvre d'Etienne-Jules Marey et sa contribution à l'émergence de la phonétique dans les sciences du langage. *In*: 2004. Disponível em: <<https://api.semanticscholar.org/CorpusID:191127851>>

TSANG, Clement Fuji *et al.* **ArtisanGS: Interactive Tools for Gaussian Splat Selection with AI and Human in the Loop.** arXiv, , 2026. Disponível em: <<https://arxiv.org/abs/2602.10173>>. Acesso em: 16 fev. 2026

TURKLE, Sherry. **Life on the screen: identity in the age of the Internet**. London: Weidenfeld & Nicolson, 1996.

ULLMAN, S. The interpretation of structure from motion. **Proceedings of the Royal Society of London. Series B. Biological Sciences**, v. 203, n. 1153, p. 405–426, 15 jan. 1979.

VELHO, Luiz *et al.* **Space XR: Ambientes Colaborativos em Realidade Expandida**. [S.l.: S.n.]. Disponível em: <https://www.visgraf.impa.br/Data/RefBib/PS_PDF/tr-08-2021/tr-08-2021.pdf>.

Venus of Willendorf. , 30 jul. 2025. (Nota técnica).

Venus of Willendorf. Disponível em: <<https://artsandculture.google.com/asset/venus-of-willendorf/EgHv9xkYFyYrIw>>. Acesso em: 22 fev. 2026.

VERMEER, Johannes Vermeer: **Senhora tocando virginal**. , 1672 1670. Disponível em: <[https://commons.wikimedia.org/wiki/File:Johannes_Vermeer_-_Zittende_Klavercimbelpeelster_\(1673-1675\).jpg?uselang=pt#Descri%C3%A7%C3%A3o_do_ficheiro](https://commons.wikimedia.org/wiki/File:Johannes_Vermeer_-_Zittende_Klavercimbelpeelster_(1673-1675).jpg?uselang=pt#Descri%C3%A7%C3%A3o_do_ficheiro)>

V-Horus – Virtual Egyptology. , [S.d.]. Disponível em: <<https://visgraf.impa.br/v-horus/>>. Acesso em: 30 mar. 2026

WANG, Nan *et al.* **Unifying Appearance Codes and Bilateral Grids for Driving Scene Gaussian Splatting**. arXiv, , 2025. Disponível em: <<https://arxiv.org/abs/2506.05280>>. Acesso em: 16 fev. 2026

Welcome to Light Fields no Steam. Disponível em: <https://store.steampowered.com/app/771310/Welcome_to_Light_Fields/>. Acesso em: 26 mar. 2026.

WU, Qi *et al.* **3DGUT: Enabling Distorted Cameras and Secondary Rays in Gaussian Splatting**. arXiv, , 24 mar. 2025. Disponível em: <<http://arxiv.org/abs/2412.12507>>. Acesso em: 19 mar. 2026

XGRIDS - 3D Spatial Intelligence Technology. Disponível em: <<https://www.xgrids.com/>>. Acesso em: 22 mar. 2026.

YANG, Zhendong *et al.* **Effective Whole-body Pose Estimation with Two-stages Distillation**. arXiv, , 2023. Disponível em: <<https://arxiv.org/abs/2307.15880>>. Acesso em: 22 mar. 2026

ZHANG, Lvmin; RAO, Anyi; AGRAWALA, Maneesh. **Adding Conditional Control to Text-to-Image Diffusion Models**. arXiv, , 2023. Disponível em: <<https://arxiv.org/abs/2302.05543>>. Acesso em: 22 mar. 2026

ZWICKER, M. *et al.* EWA volume splatting. *In: VIS 2001. VISUALIZATION 2001. Proceedings Visualization, 2001. VIS '01.* San Diego, CA, USA: IEEE, 2001. Disponível em: <<http://ieeexplore.ieee.org/document/964490/>>. Acesso em: 4 fev. 2026