



# CAPÍTULO 12

## TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL APLICADAS À ANIMAÇÃO FACIAL DE PERSONAGENS VIRTUAIS

**Carlos Eduardo Rossi Cubas da Silva**

Doutor em Mídia e Tecnologia pela Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Arquitetura, Artes, Comunicação e Design - Câmpus de Bauru e mestre em Ciência da Computação pela UNESP.

**Antonio Carlos Sementille**

Docente e pesquisador do Programa de Pós-Graduação em Mídia e Tecnologia (Doutorado) na Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Arquitetura, Artes, Comunicação e Design - Câmpus de Bauru.

### RESUMO

Nas últimas décadas, o interesse quanto à captura de movimentos da face humana e à identificação de suas expressões, com a finalidade de geração de animações faciais realistas, tem aumentado, tanto na comunidade científica quanto na indústria do entretenimento. A alta acurácia nesse processo é necessária, pois os humanos são naturalmente treinados para identificar expressões faciais, detectando facilmente pequenas imperfeições na animação de uma face virtual. O redirecionamento de movimento facial é um dos métodos utilizados para gerar animações realistas, principalmente em filmes. Muitos sistemas foram desenvolvidos e consistem, de maneira geral, em uma fase de modelagem, captura de movimentos e identificação de expressões faciais, seguido de um procedimento de redirecionamento de expressões para uma face virtual. Para o redirecionamento são aplicadas várias técnicas, onde algoritmos de cálculo de distância ou redes neurais são utilizados, fazendo com que a face final gerada no processo seja próxima a da captura. Nesse sentido, a utilização de técnicas de inteligência artificial é extremamente útil, permitindo que a acurácia e o desempenho do processo sejam melhoradas. Considerando este contexto, o objetivo principal deste capítulo é apresentar um panorama geral do problema da animação facial de personagens, bem como, descrever brevemente alguns dos principais trabalhos do estado da arte nessa área, com especial ênfase naqueles que utilizam técnicas de Inteligência Artificial.

**Palavras-chave:** animação facial; *blendshapes*; *retargeting*; inteligência artificial.

### 1 INTRODUÇÃO

Nas últimas décadas, o interesse quanto à captura de movimentos da face e a identificação de suas expressões, com a finalidade de gerar animações faciais realistas, em personagens virtuais 2D e 3D, tem aumentado, tanto na comunidade científica quanto na indústria do entretenimento. Com uma animação facial de qualidade, é possível fazer com que personagens virtuais transmitam emoções quase reais. No entanto, esse tipo de animação é uma tarefa complexa, pois lida com inúmeras variações de movimento do rosto.

A grande complexidade e sofisticação da estrutura da cabeça humana aumenta a dificuldade da reprodução de uma animação facial convincente. Uma alta acurácia é necessária, porque os humanos são treinados para observar e decodificar expressões faciais desde o



nascimento, tornando-os especialistas na detecção de pequenos erros na animação de uma face virtual (FRATARCANGELI, 2013).

O cinema é um dos grandes beneficiados com a captura de movimentos faciais. O personagem Gollum da franquia Senhor dos Anéis (Figura 1) e os filmes King Kong e Avatar, são exemplos de produções cinematográficas que utilizaram esse tipo de animação (SEOL et al., 2011).

**Figura 1** – Personagem Gollum da franquia Senhor dos Anéis.



**Fonte:** Gelman, (2020).

Já, no mundo dos jogos, essa tecnologia pode ser usada para controlar avatares, representando a face do usuário, permitindo que os seus amigos reais percebam as expressões do usuário refletidas no mundo virtual, em tempo real.

A medicina, também pode se beneficiar dessa tecnologia, nas quais novas formas de terapia interativa são possíveis. Por exemplo: sessões de treinamento baseadas em avatares podem ser criadas para pessoas com autismo ou outros transtornos do desenvolvimento neural (PAULY, 2013).

Um ambiente imersivo onde esta tecnologia também pode ser aplicada é o metaverso que, segundo Lee et al. (2021), é um ambiente virtual que mistura o físico e o digital onde os usuários possuem seus respectivos avatares e experimentam uma alternativa à vida física em uma realidade virtual. De acordo com Chen et al. (2022), o metaverso é um espaço de vida e ciberespaço que realiza o processo de virtualização e digitalização do mundo real.

Com isso, o uso de avatares no metaverso torna-se uma necessidade, proporcionando aos usuários a experiência de viver em um mundo paralelo virtual, onde ele poderá se comunicar com outros avatares através de texto, voz e gestos, criando uma sociedade totalmente virtual. Desta forma, o uso das tecnologias de captura dos movimentos faciais no metaverso, podem



proporcionar, além da comunicação verbal, uma comunicação através das mudanças de expressões nas faces virtuais dos avatares em tempo real, proporcionando uma experiência mais imersiva e mais próxima da humana.

O redirecionamento de movimento facial é uma das técnicas utilizadas para a geração de animações faciais. Ele é composto por três principais etapas: modelagem, captura e redirecionamento (*retargeting*). Cada etapa é responsável pelo processamento de várias informações, como a geração de um banco de dados de expressões, captura da face do ator em tempo real, cálculo para gerar a nova face e transferência para uma face virtual (SHAKIR; AL-AZZA, 2022).

Na etapa da modelagem são construídos modelos digitais de faces que podem ser armazenadas, exibidas e modificadas digitalmente. Duas abordagens se destacam: a deformação de malha e os *blendshapes*.

Na abordagem da deformação de malha, uma única malha é deformada ao longo da captura, como mostrado na Figura 2.

Na abordagem dos *blendshapes*, por sua vez, utilizam-se modelos faciais, onde vetores representam expressões individuais, para representar as expressões (Figura 3).

Para a geração de uma nova face virtual são utilizadas combinações de vários *blendshapes*, por meio do cálculo da relevância das expressões armazenadas em uma base de dados, com relação à face real capturada. Esse cálculo resulta em pesos, valores que serão multiplicados pelas expressões armazenadas e somadas a uma das expressões base, gerando a face final.

**Figura 2** – Exemplo da deformação de malha.



**Fonte:** Wu *et al.* (2016).



**Figura 3** – Exemplo de blendshapes.

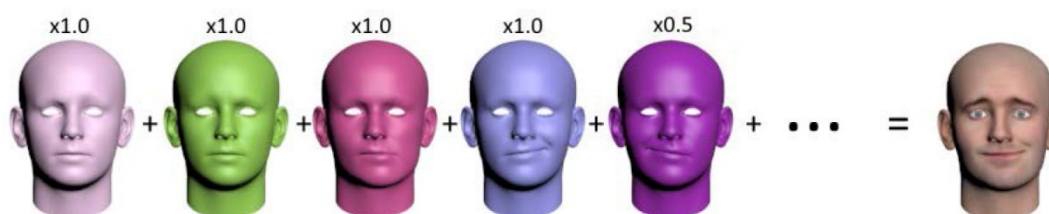


**Fonte:** Lewis et al. (2014).

Ao variar as combinações, através de uma soma de vetores, uma gama de expressões faciais pode ser criada com pequeno esforço computacional (WEISE et al., 2011). Um exemplo pode ser visto na Figura 4, onde vários *blendshapes* são somados, cada um com uma relevância, resultando, no final, em uma nova face.

Na etapa de captura, existem vários métodos que podem ser usados para realizar o redirecionamento de movimento facial. Um método muito usado é o método óptico, onde câmeras especiais são colocadas à frente do rosto do ator e marcadores com tinta refletiva são pintados em pontos estratégicos de sua face (Figura 5).

**Figura 4** – Criação de uma face utilizando a combinação de *blendshapes*.



**Fonte:** Costigan et al. (2016).

Porém, a transposição desses movimentos para a face de um personagem virtual não é trivial e, geralmente, exige a execução de vários procedimentos durante a fase de pós-produção, sendo um processo quase artesanal (CUBAS; SEMENTILLE, 2017).

**Figura 5** – Exemplo da utilização de marcadores com tinta refletiva.



**Fonte:** Pennington (2018).

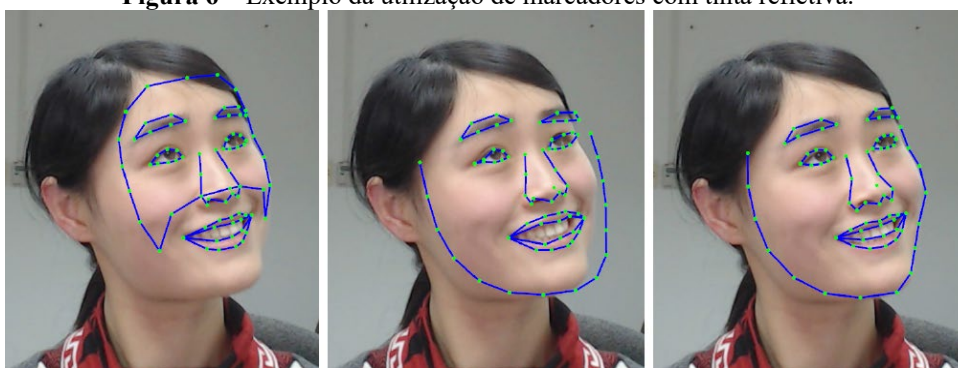




Outro método utiliza a captura de forma automática, através de câmeras de captura de vídeo, como as *webcams*. Elas registram a face do ator e capturam os marcadores faciais, que são informações de pontos específicos, como o contorno da boca, dos olhos, do nariz, das sobrancelhas, entre outros, que servirão de referências para a construção da nova face (Figura 6).

Nessa etapa, alguns trabalhos como os de Weise et al. (2011) e Li et al. (2013) utilizam câmeras especiais com informações RGB-D (*Red, Green, Blue - Depth*), que capturam, além das imagens 2D, a informação de profundidade associada a cada *pixel*. Outros trabalhos, como os de Saito, Li e Li (2016) e Cao, Hou e Zhou (2014), utilizam informações de câmeras RGB que contêm apenas informações 2D.

**Figura 6** – Exemplo da utilização de marcadores com tinta refletiva.



Fonte: Cao, Hou e Zhou (2014).

Na etapa de redirecionamento, os parâmetros obtidos na etapa de captura são utilizados para animar os personagens virtuais. Essa tarefa é altamente não trivial, especialmente, quando o personagem alvo não é uma réplica próxima da face capturada e tem proporções diferentes dela.

Existem várias abordagens para o redirecionamento, como a utilização do PCA (*Principal Component Analysis*), AAM (*Active Appearance Model*) (ZHOU et al, 2010) ou redes neurais como a CNN (*Convolutional Neural Networks*) (GUO et al., 2017).

## 2 TRABALHOS RELACIONADOS

Nesta seção são apresentados os principais trabalhos encontrados na literatura científica, que usam técnicas de Inteligência Artificial para a animação facial, especialmente, na etapa de redirecionamento dos movimentos faciais. Uma especial ênfase é dada aos trabalhos que se baseiam na abordagem dos *blendshapes*, na etapa de modelagem das faces virtuais.

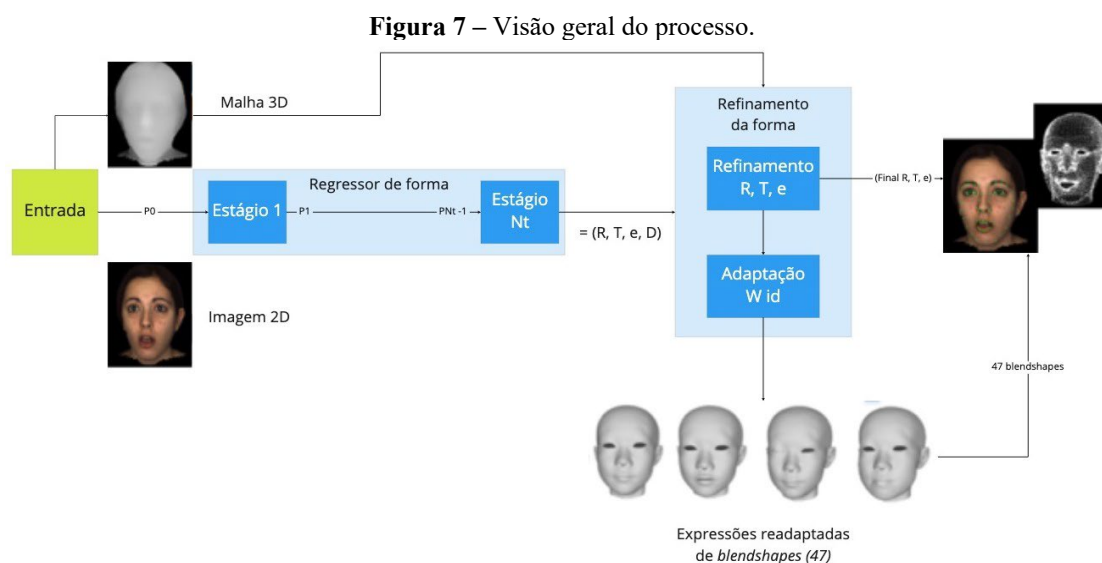


## 2.1 MÉTODO DE PHAM et al. (2016)

Pham et al. (2016) propôs um método para redirecionamento da animação facial, que combina as vantagens da regressão de imagens RGB e os pontos 3D capturados por uma câmera RGB-D. Ele usa o algoritmo de Floresta Aleatória para treinar uma rede que calcula os pesos a serem aplicados aos *blendshapes* (HO, 1995).

Esse método traz inovação com possibilidade de que as capturas para o redirecionamento sejam feitas em uma distância maior do que em outros trabalhos. Dessa maneira, esse método foca em problemas onde o ator não está próximo como, por exemplo, uma sala de videoconferência.

O sistema proposto consiste em dois módulos principais: o módulo de regressão e o módulo de refinamento. O módulo de regressão, executa o primeiro estágio de otimização das características, através de uma rede treinada com as informações 2D capturadas da imagem de uma base de dados a *Face Warehouse* (CAO et al., 2014). O refinamento utiliza, além dos dados 2D, os dados 3D da nuvem de pontos encontrados nas imagens RGB-D (Figura 7).



Fonte: Pham et al. (2016).

Para a geração dos pesos, a rede faz uma predição da possibilidade da captura pertencer a alguma expressão treinada. Isso faz com que seja gerado um vetor de características com os valores normalizados entre 0 e 1. Este vetor é usado para a construção da face final.

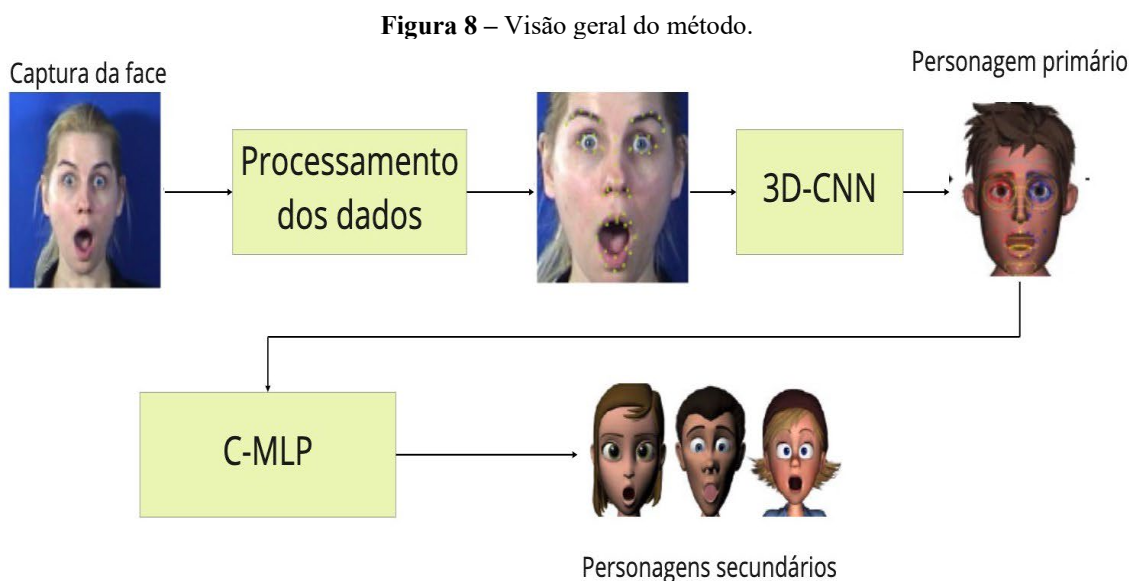
## 2.2 MÉTODO DE ANEJA et al. (2018)

Aneja et al. (2018) desenvolveram um método semi-supervisionado de redirecionamento de movimento facial que, através do mapeamento 2D de um rosto humano, prevê os parâmetros necessários para animação da face de personagens 3D estilizados.



O sistema que foi proposto utiliza bases de dados de expressões faciais divididas em seis categorias: raiva, nojo, medo, alegria, tristeza, surpresa e a expressão neutra. Essas expressões ocorrem em uma ampla gama de intensidade e podem se misturar para criar expressões adicionais.

Através da análise de similaridade, é treinada uma rede neural convolucional CNN (INDOLIA et al., 2018) que faz a análise características 2D e transferência de expressões da face do ator para o personagem 3D (vide Figura 8).



**Fonte:** Aneja et al. (2018).

Na etapa de captura da face, são extraídos 49 marcadores faciais. Os marcadores são usados como pontos de referência para extrair as características geométricas da face, incluindo a altura da sobrancelha esquerda, altura da sobrancelha direita, a distância entre o topo da sobrancelha, o centro do olho, altura da pálpebra esquerda, altura da pálpebra direita, largura do nariz, medida de boca fechada, altura do lábio esquerdo e altura do lábio direito. As distâncias das áreas analisadas são normalizadas pelos marcadores faciais do contorno da face.

Na primeira etapa é feita a captura da face do ator. Após a captura, é feito o processamento por meio da extração dos marcadores faciais. Os marcadores faciais obtidos na etapa de captura são utilizados para rotular uma base de dados de personagens da FERG-DB (ANEJA et al., 2016), para encontrar semelhança entre eles, através da distância euclidiana (ROBISON, 2014). Com isso a base é rotulada.

Após essa etapa, uma rede do tipo CNN é treinada para detectar a similaridade da captura da face com os personagens. No final, uma rede do tipo MLP, *Multi-Layer Perceptron* (CHOI; PARK, 2021) é utilizada para fazer a transferência da expressão para o personagem.



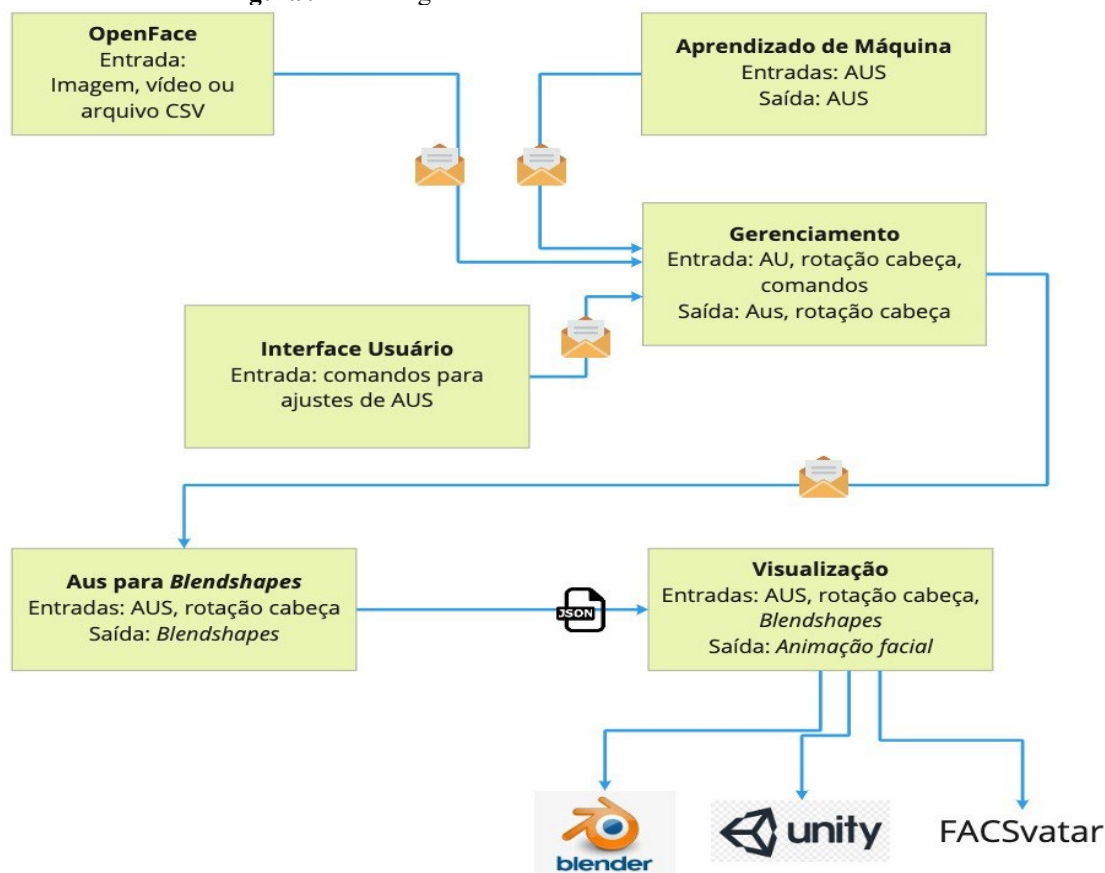
## 2.3 MÉTODO DE STRUIJK et al. (2018)

Struijk et al. (2018) desenvolveram um *framework* chamado *FACSVatar* criado para possibilitar a animação facial baseado nas FACS (*Facial Action Coding System*) (EKMAN; ROSEMBERG, 1997) em *softwares* como *Unity 3D* ou *Blender*. Ele utiliza implementações de código aberto para compor os módulos propostos.

Todos os módulos do *framework* se comunicam de forma assíncrona, trocando mensagens entre si (Figura 9). Os módulos efetivados e testados foram:

- Entrada dos FACS, movimento dos olhos e rotação da cabeça em tempo real e *offline*;
- Gerenciamentos das mensagens assíncronas;
- Interface gráfica para ajustes das AUS (EKMAN; ROSEMBERG, 1997);
- Rede Neural para geração da animação facial;
- Conversão dos AUS para os *blendshapes*; e
- Visualização dos resultados *Blender* (COMMUNITY, 2018), *Unity3D* (HAAS, 2014) ou *FACSHuman* (GILBERT; DEMARCHI; URDAPILLETA, 2018).

Figura 9 – Visão geral da estrutura de módulos desenvolvido.



Fonte: Struijk et al. (2018).



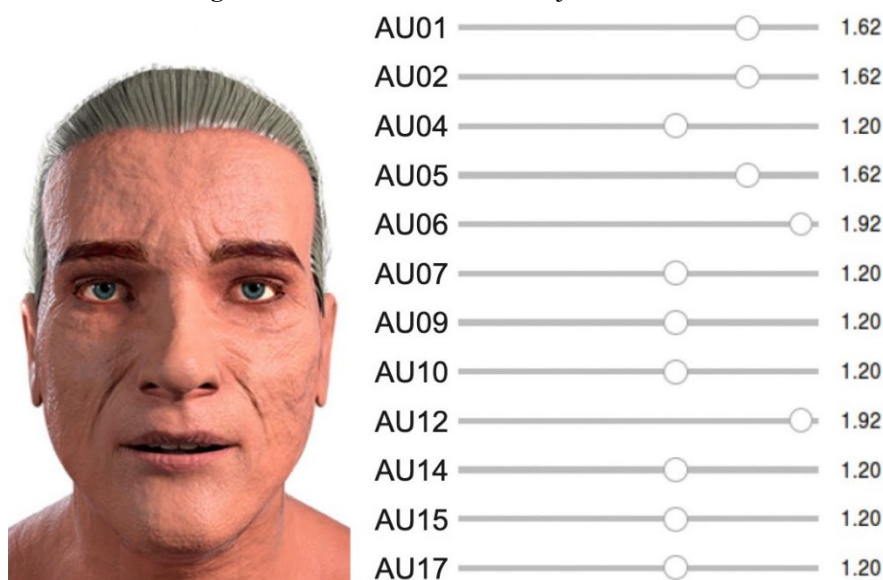


Para o módulo inicial foi utilizada a implementação *OpenFace 2.0*, uma ferramenta para pesquisadores de visão computacional e aprendizado de máquina e pessoas interessadas em construir aplicativos interativos baseados em análise de comportamento facial. Segundo Baltrusaitis et al. (2018), o módulo foi modificado para enviar os AU da face capturada, a posição dos olhos e a rotação da cabeça em tempo real.

O controle de informações entre os módulos é feito de maneira assíncrona, por meio da troca de mensagens. Foi desenvolvido um gerenciador de mensagens que funciona como uma central de envio e recebimento de informações, que são repassadas aos outros módulos do *pipeline*.

A interface gráfica para ajuste das AUS foi desenvolvida em um módulo separado, através de comandos horizontais, que representam os tipos de AUS (Figura 10).

**Figura 10** – Interface do módulo ajuste dos AUS.



**Fonte:** Struijk et al. (2018).

O módulo de aprendizado de máquina utilizou uma rede neural GRU (*Gated Recurrent Unit*) (CHO et al., 2014) implementada em Keras (CHOLLET et al., 2015), que calcula a animação facial através da análise das AUS. A saída se dá, por meio, dos valores de 17 diferentes AUS indicando a sua relevância para o redirecionamento. Os dados da base de dados MAHNOB Mimicry (BILAKHIA et al., 2015) foram usados para o treinamento.

No módulo de conversão dos AUS para os *blendshapes*, é feita uma análise de todas as possibilidades disponíveis, encontrando assim, a melhor combinação entre os *blendshapes* para as expressões identificadas pelos AUS.



Como o *framework* foi desenvolvido para ser um *plugin*, o resultado final se dá nos softwares *Unity 3D*, *FACSHuman* ou *Blender*.

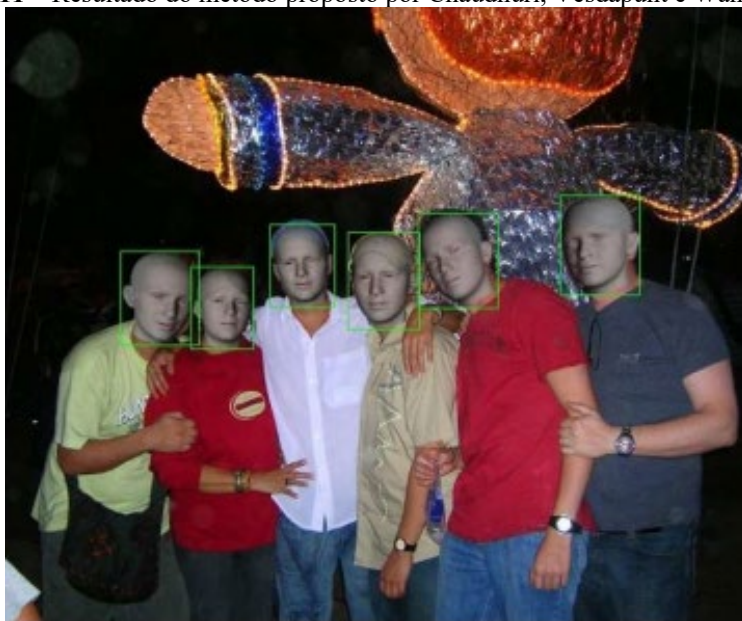
## 2.4 MÉTODO DE CHAUDHURI, VESDAPUNT e WANG (2019)

Chaudhuri, Vesdapunt e Wang (2019) propuseram um método que prevê o redirecionamento da animação facial para múltiplas faces, diferente de outros métodos que trabalham com uma única face, através de uma rede neural convolucional CNN. Para treinar essa rede são usados os parâmetros da 3DMM (BLANZ; VETTER, 1999) para a extração das características.

O método é dividido em duas partes. Na primeira parte, é treinada uma rede utilizando os parâmetros 3DMM para detecção de uma única face chamada SFN (*Single Face Network*). São eles: parâmetros de formato da face, parâmetros que capturam a expressão facial, parâmetros da rotação 3D da cabeça e parâmetros de escala que vinculam a face 3D à imagem 2D. Na segunda parte, o processo é expandido para detectar múltiplas faces em uma única imagem. Um exemplo do método pode ser visto na Figura 11.

Para o redirecionamento da animação facial final, a rede treinada gera os pesos finais baseados nas entradas extraídas dos parâmetros 3DMM.

**Figura 11** – Resultado do método proposto por Chaudhuri, Vesdapunt e Wang (2019).



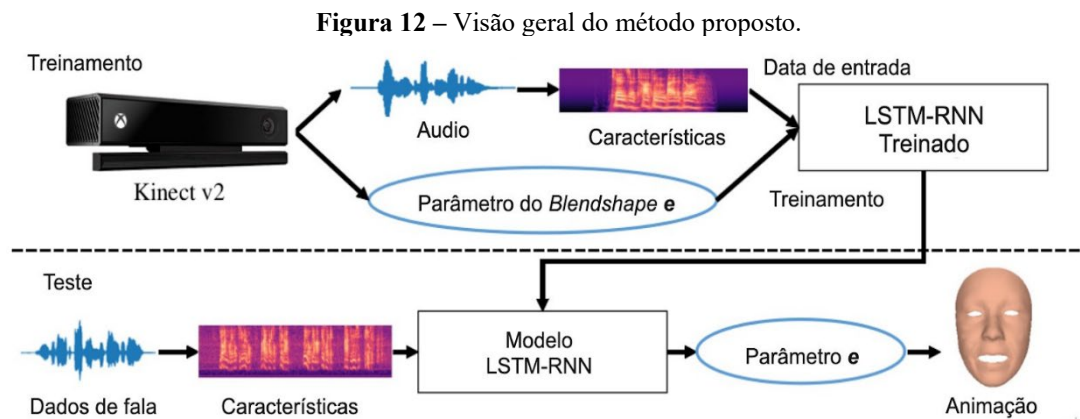
Fonte: Chaudhuri, Vesdapunt e Wang (2019).

## 2.5 MÉTODO DE NISHIMURA et al. (2019)

Nishimura et al. (2019) apresentaram um método para redirecionamento da animação facial que utiliza a fala. Segundo os autores, os modelos tradicionais que usam uma câmera



convencional podem causar limitações como, por exemplo, o ângulo da captura que, dependendo da posição da face, pode comprometer o processo. Uma visão geral do método pode ser vista na Figura 12.



Fonte: Nishimura et al. (2019).

Nesse método, é utilizado LSTM (*Long Short Term Memory*), que é uma variação da RNN (*Recurrent Neural Networks*) (ACADEMY, 2019) onde a entrada são os dados de análise da fala e a saída os pesos para os *blendshapes*. O sistema usa um dispositivo RGBD, o *Kinect*, para a captura dos pontos da face e geração dos *blendshapes*.

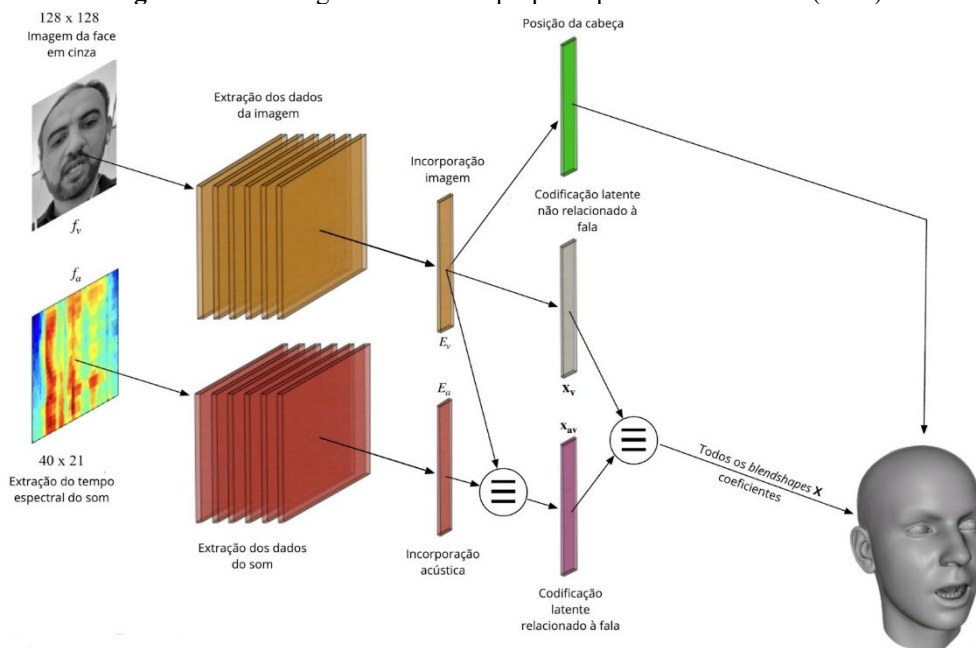
O sistema proposto estima expressões faciais de fala por LSTM-RNN, porque os dados de áudio são um sinal com séries temporais possuindo uma característica que se propaga, sequencialmente, ao longo do tempo. A rede LSTM-RNN, através dos dados de áudio, gera os pesos dos *blendshapes* em valores entre 0 e 1.

## 2.6 MÉTODO DE ABDELAZIZ et al. (2020)

O método de redirecionamento facial proposto por Abdelaziz et al. (2020) utiliza como entrada de dados informações acústicas e visuais. O modelo proposto roda em tempo real, em *hardware* com recursos limitados como, por exemplo, um telefone celular e é independente, podendo funcionar com qualquer tipo de usuário. A saída são pesos para os *blendshapes* baseados nas FACS (EKMAN; ROSEMBERG, 1997). A Figura 13 apresenta uma visão geral do *pipeline* para rastreamento e animação.



Figura 13 – Visão geral do método proposto por Abdelaziz et al. (2020).



Fonte: Abdelaziz et al. (2020).

Os recursos visuais de entrada  $f_v$  são imagens em escala de cinza. As faces são detectadas, através de uma rede neural pré-treinada quadro a quadro. Os recursos acústicos de entrada  $f_a$  são extraídos das amostras de fala bruta. São capturados, simultaneamente e, no mesmo dispositivo, informações de áudio e vídeo, para que os sinais sejam sincronizados no momento da captura.

Para o cálculo dos pesos utilizados na geração da face final, o método proposto treinou uma DNN (*Deep neural networks*) (GAWLIKOWSKI et al, 2021), para mapear imagens da face de entrada para coeficientes de *blendshapes*. A rotulação da base de dados de treinamento da rede foi feita através de um método proposto por Weise et al. (2011), que gerou os coeficientes, a partir de uma série de modelos personalizados. Esses modelos foram criados adaptando um modelo genérico de *blendshape*.

### 3 CONSIDERAÇÕES FINAIS

O quadro 1 sintetiza as principais características dos métodos levantados na seção anterior.

Quadro 1 – Principais características dos métodos levantados.

Método	Captura	Algoritmo	Geração de pesos
Pham et al. (2016)	RGB e RGB-D	RF	Dados 3D e 2D
Aneja et al. (2018)	RGB	CNN	Dados 2D
Struijk et al. (2018)	RGB	GRU	Dados 2D



Chaudhuri, Vesdapunt e Wang (2019)	RGB	CNN	Parâmetros 3DMM
Nishimura et al. (2019)	Som	LSTM	Fala
Abdelaziz et al. (2020)	Som e Vídeo	DNN	Dados 3D e 2D

Fonte: elaborada pelo autor.

O método apresentado por Pham et al. (2016) propôs o redirecionamento de faces a longas distâncias, mas o processo possui configurações em duas etapas dependentes uma da outra, dificultando a configuração inicial para o redirecionamento.

Aneja et al. (2018) apresentou um método de animação facial para personagens estilizados. São utilizados marcadores faciais para identificar as características e treinamento da base de dados, de forma não supervisionada. Após a identificação dos personagens, é treinada uma rede do tipo CNN para a transferência das expressões entre a face capturada e o personagem.

O trabalho apresentado por Struijk et al. (2018) mostra o desenvolvimento de um framework para redirecionamento facial. A ideia principal foi o desenvolvimento de um modelo flexível, com módulos se comunicando de forma assíncrona e utilizando projetos de código aberto. Implementado na forma de um plugin, sua saída pode ser redirecionada para os softwares *Blender*, *Unity3D* ou *FACSHuman*.

Chaudhuri, Vesdapunt e Wang (2019) também apresentam um método inovador prevendo o redirecionamento facial para várias capturas na mesma imagem, mas o processo de treinamento para geração dos pesos passa por duas etapas, sendo a primeira, para uma única foto e a segunda, para várias fotos, levando em conta algumas informações do primeiro processo. Como são processos complementares, eles compartilham informações entre si, dificultando a calibração inicial e o redirecionamento da face.

O método de Nishimura et al. (2019) sugere o redirecionamento facial através da captura da fala, visto que, alguns métodos mais tradicionais que utilizam a captura da face, podem sofrer problemas quando as obstruções como por exemplo, com o cabelo na frente dos olhos ou sobrancelhas e a posição da cabeça. Para melhorar, esse método aponta problemas na sincronização das expressões, principalmente, no movimento da boca.

Já o método proposto por Abdelaziz et al. (2020) utiliza a fala e imagens sincronizadas e uma rede DNN para o cálculo dos pesos dos blendshapes.

A partir deste levantamento, sintetizado no Quadro 1, nota-se a crescente utilização das técnicas de Inteligência Artificial mais sofisticadas, convergindo-se atualmente para as redes





neurais mais profundas, uma vez que, além das características inerentes às imagens, outras ligadas, por exemplo, ao áudio sincronizado, também estão sendo utilizadas.

## REFERÊNCIAS

ABDELAZIZ, A. H.; THEOBALD, B.-J.; DIXON, P.; KNOTHE, R.; APOSTOLOFF, N.; KAJAREKER, S. **Modality dropout for improved performance-driven talking faces**. In: Proceedings of the 2020 International Conference on Multimodal Interaction. New York, NY, USA: Association for Computing Machinery, 2020. p. 378–386. ISBN 9781450375818.

ACADEMY, D. S. *Arquitetura de Redes Neurais Long Short Term Memory (LSTM)*. 2019. Disponível em: <<http://deeplearningbook.com.br/arquitetura-de-redes-neurais-long-short-term-memory>>.

ANEJA, D.; CHAUDHURI, B.; COLBURN, A.; FAIGIN, G.; SHAPIRO, L.; MONES, B. **Learning to generate 3d stylized character expressions from humans**. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). [S.l.: s.n.], 2018. p. 160–169.

ANEJA, D.; COLBURN, A.; FAIGIN, G.; SHAPIRO, L.; MONES, B. **Modeling stylized character expressions via deep learning**. In: SPRINGER. Asian Conference on Computer Vision. [S.l.], 2016. p. 136–153.

BALTRUSAITIS, T.; ZADEH, A.; LIM, Y. C.; MORENCY, L.-P. Openface 2.0: **Facial behavior analysis toolkit**. In: 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018). [S.l.: s.n.], 2018. p. 59–66.

BILAKHIA, S.; PETRIDIS, S.; NIJHOLT, A.; PANTIC, M. **The mahnob mimicry database: A database of naturalistic human interactions**. *Pattern Recognition Letters*, v. 66, p. 52–61, 2015. ISSN 0167-8655. Pattern Recognition in Human Computer Interaction. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865515000768>>.

BISHOP, C. M. **Pattern recognition and machine learning**. [S.l.]: springer, 2006.

BLANZ, V.; VETTER, T. **A morphable model for the synthesis of 3d faces**. In: ACM PRESS/ADDISON-WESLEY PUBLISHING CO. Proceedings of the 26th annual conference on Computer graphics and interactive techniques. [S.l.], 1999. p. 187–194.

CAO, C.; HOU, Q.; ZHOU, K. **Displaced dynamic expression regression for real-time facial tracking and animation**. *ACM Trans. Graph.*, Association for Computing Machinery, New York, NY, USA, v. 33, n. 4, jul 2014. ISSN 0730-0301. Disponível em: <<https://doi.org/10.1145/2601097.2601204>>.

CAO, C.; WENG, Y.; ZHOU, S.; TONG, Y.; ZHOU, K. **Facewarehouse: A 3d facial expression database for visual computing**. *Visualization and Computer Graphics*, IEEE Transactions on, v. 20, n. 3, p. 413–425, March 2014. ISSN 1077-2626.

CHAUDHURI, B.; VESDAPUNT, N.; WANG, B. **Joint face detection and facial motion retargeting for multiple faces**. In: [S.l.: s.n.], 2019. p. 9711–9720.

CHEN, Z.; WU, J.; GAN, W.; QI, Z. **Metaverse Security and Privacy: An Overview**. arXiv, 2022. Disponível em: <<https://arxiv.org/abs/2211.14948>>.



CHO, K.; MERRIËNBOER, B. V.; GULCEHRE, C.; BAHDANAU, D.; BOUGARES, F.; SCHWENK, H.; BENGIO, Y. **Learning phrase representations using rnn encoder-decoder for statistical machine translation**. arXiv preprint arXiv:1406.1078, 2014.

CHOI, H.; PARK, S. **A survey of machine learning-based system performance optimization techniques**. Applied Sciences, Multidisciplinary Digital Publishing Institute, v. 11, n. 7, p. 3235, 2021.

CHOLLET, F. et al. **Keras**. GitHub, 2015. Disponível em: <<https://github.com/fchollet/keras>>.

COMMUNITY, B. O. **Blender - a 3D modelling and rendering package**. Stichting Blender Foundation, Amsterdam, 2018. Disponível em: <<http://www.blender.org>>.

COSTIGAN, T.; GERDELAN, A.; CARRIGAN, E.; MCDONNELL, R. **Improving blendshape performance for crowds with gpu and gpgpu techniques**. In: Proceedings of the 9th International Conference on Motion in Games. New York, NY, USA: Association for Computing Machinery, 2016. (MIG '16), p. 73–78. ISBN 9781450345927.

CUBAS, C.; SEMENTILLE, A. **The development of a facial animation system based on performance and the use of an rgb-d camera**. In: 2017 19th Symposium on Virtual and Augmented Reality (SVR). [S.l.: s.n.], 2017. p. 78–86.

EKMAN, P.; ROSENBERG, E. L. **What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)**. [S.l.]: Oxford University Press, 1997.

FRATARCANGELI, M. **Computational models for animating 3d virtual faces**. Linköping University Electronic Press, 2013.

GAWLIKOWSKI, J.; TASSI, C. R. N.; ALI, M.; LEE, J.; HUMT, M.; FENG, J.; KRUSPE, A.; TRIEBEL, R.; JUNG, P.; ROSCHER, R. et al. **A survey of uncertainty in deep neural networks**. arXiv preprint arXiv:2107.03342, 2021.

GELMAN, S. First Concept Art from Lord of the Rings: **Gollum Game Surfaces**. 2020. <<https://bityli.com/Wk28Vv>>. Accessed: 2022-09-26.

GILBERT, M.; DEMARCHI, S.; URDAPILLETA, I. **Facshuman, a software to create experimental material by modeling 3d facial expression**. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. New York, NY, USA: Association for Computing Machinery, 2018. (IVA '18), p. 333–334. ISBN 9781450360135.

GUO, Y.; ZHANG, J.; CAI, J.; JIANG, B.; ZHENG, J. **Cnn-based real-time dense face reconstruction with inverse-rendered photo-realistic face images**. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 41, p. 1294–1307, 2017.

HAAS, J. K. **A history of the unity game engine**. Worcester Polytechnic Institute, 2014.

HO, T. K. **Random decision forests**. In: Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1. M: IEEE Computer Society, 1995. (ICDAR '95), p. 278–282. ISBN 0-8186-7128-9.



INDOLIA, S.; GOSWAMI, A.; MISHRA, S.; ASOPA, P. **Conceptual understanding of convolutional neural network - a deep learning approach.** *Procedia Computer Science*, v. 132, p. 679–688, 01 2018.

LEE, L.-H.; BRAUD, T.; ZHOU, P.; WANG, L.; XU, D.; LIN, Z.; KUMAR, A.; BERMEJO, C.; HUI, P. **All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda.** *arXiv*, 2021.

LEWIS, J. P.; ANJYO, K.; RHEE, T.; ZHANG, M.; PIGHIN, F. H.; DENG, Z. **Practice and theory of blendshape facial models.** In: CITESEER. *Eurographics (State of the Art Reports)*. [S.l.], 2014. p. 199–218.

LI, H.; YU, J.; YE, Y.; BREGLER, C. **Realtime facial animation with on-the-fly correctives.** *ACM Trans. Graph.*, ACM, New York, NY, USA, v. 32, n. 4, p. 42:1–42:10, jul. 2013. ISSN 0730-0301.

NISHIMURA, R.; SAKATA, N.; TOMINAGA, T.; HIJIKATA, Y.; HARADA, K.; KIYOKAWA, K. **Speech-driven facial animation by lstm-rnn for communication use.** 2019 12th Asia Pacific Workshop on Mixed and Augmented Reality (APMAR), p. 1–8, 2019.

PANCERI, J. A. **Reconhecimento de Expressões Faciais Baseado em Active Appearance Model.** 2017.

PAULY, M. **Realtime performance-based facial avatars for immersive gameplay.** In: ACM. *Proceedings of Motion on Games*. [S.l.], 2013. p. 23–28.

PENNINGTON, A. **What Do Motion Capture Actors Actually Do?** 2018. <<https://bityli.com/y0U9g3>>. Accessed: 2022-09-26.

Pham, H. X.; Pavlovic, V.; Jianfei Cai; Tat-jen Cham. **Robust real-time performance-driven 3d face tracking.** In: 2016 23rd International Conference on Pattern Recognition (ICPR). [S.l.: s.n.], 2016. p. 1851–1856.

ROBISON, S. **Geometria Plana.** 2014. <<https://bityli.com/DF2rqs>>. Accessed: 2016-12-05.

SAITO, S.; LI, T.; LI, H. **Real-time facial segmentation and performance capture from rgb input.** *ArXiv*, abs/1604.02647, 2016.

SEOL, Y.; SEO, J.; KIM, P. H.; LEWIS, J. P.; NOH, J. **Artist friendly facial animation retargeting.** *ACM Trans. Graph.*, ACM, New York, NY, USA, v. 30, n. 6, p. 162:1–162:10, dez. 2011. ISSN 0730-0301.

SHAKIR, S. D.; AL-AZZA, A. A. **Facial modeling and animation: An overview of the state-of-the art.** *Iraqi Journal for Electrical & Electronic Engineering*, v. 18, n. 1, 2022.

STATISTICS, L. B.; BREIMAN, L. Random forests. In: *Machine Learning*. [S.l.: s.n.], 2001. p. 5–32.

STRUIJK, S. van der; HUANG, H.-H.; MIRZAEI, M. S.; NISHIDA, T. **Facsvatar: An open source modular framework for real-time faces based facial animation.** In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. New York, NY, USA: Association for Computing Machinery, 2018. (IVA '18), p. 159–164. ISBN 9781450360135.



WEISE, T.; BOUAZIZ, S.; LI, H.; PAULY, M. **Realtime performance-based facial animation**. In: ACM. ACM Transactions on Graphics (TOG). [S.l.], 2011. v. 30, n. 4, p. 77.

WU, C.; BRADLEY, D.; GROSS, M.; BEELER, T. **An anatomically-constrained local deformation model for monocular face capture**. ACM Trans. Graph., Association for Computing Machinery, New York, NY, USA, v. 35, n. 4, jul 2016. ISSN 0730-0301.

YANGA, Y.; WU, X.-J.; KITTLER, J. **Landmark weighting for 3dmm shape fitting**. ArXiv, abs/1808.05399,2018.

ZHOU, M.; LIANG, L.; SUN, J.; WANG, Y. **Aam based face tracking with temporal matching and face segmentation**. In: IEEE. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. [S.l.], 2010. p. 701–708.