Síntese de expressões faciais baseada em faces similares

Rafael Luiz Testa

I. Trabalhos correlatos

A síntese de expressões faciais consiste em alterar a expressão facial de uma imagem (face destino neutra - D_n) com base em exemplos (face fonte expressiva - F_e) para representar uma emoção diferente, enquanto preserva as características faciais relativas à identidade da D_n. Esse processo resulta em uma imagem sintética (face destino sintética - D_s) que possui a expressão facial da F_e e a identidade da D_n. As abordagens encontradas na literatura podem ser divididas em dois grupos: síntese baseadas em Aprendizado de Máquina e Deformação-e-Mapeamento [1]. Enquanto o primeiro grupo utiliza todas as imagens disponíveis para construção do modelo, o segundo utiliza como fonte um par de imagens: uma imagem contendo a expressão facial desejada (F_e) e outra do mesmo indivíduo sem a expressão (Face fonte neutra - F_n).

Abordagens baseadas em Aprendizado de Máquina podem utilizar as imagens, os *landmarks*, e/ou características relativas a iluminação como base para construção do modelo [1]. Alguns estudos inclusive decompõem e reconstroem as expressões faciais dividindo-as em dois sub-espaços: identidade facial e expressão facial. Nesta abordagem o modelo geralmente é construído com base em técnicas de aprendizado supervisionado como regressão [2], Máquina de Vetores de Suporte [3] ou Rede Neurais [4], [5]. Estudos mais recentes têm explorado o emprego de Redes Neurais Adversativas (*Generative Adversarial Network - GANs*) para síntese da nova imagem, como, por exemplo, *Conditional Difference Adversarial Autoencoder (CDAAE)* [6], *StarGAN* [7], *Geometry Guided Adversarial Network* [8], *Warp-guided GANs* [9] and *ExprGAN* [10].

As abordagens baseadas em Deformação e Mapeamento consistem em estabelecer uma correspondência entre os componentes faciais, deformar tais componentes e mapear as diferenças de iluminação. A correspondência entre os componentes faciais em duas imagens diferentes é realizada por meio da detecção da face, identificação dos *landmarks* faciais e a triangulação desses pontos. Estas sub-etapas podem ser realizadas, respectivamente, com Rede Neurais Convolutivas [11], [12], Árvores de Regressão em Cascata [13] e triangulação de *Delaunay* [14], [15], [1]. Na deformação uma configuração facial expressiva é calculada e os *pixels* dos triângulos da face neutra são mapeados para o novo formato por meio de transformações afins [1], [14], [15]. Por fim, os aspectos finos da expressão facial, como as rugas, são mapeados utilizando *Expression Ratio Images* [1], [16], [14].

Também há abordagens que combinam passos das abordagens de síntese baseadas em Deformação-e-Mapeamento e Aprendizado de Máquina [1]. Por exemplo, uma imagem inicial é sintetizada pela primeira abordagem e combinada com a imagem gerada pela segunda abordagem. É o caso dos *Warp-guided GANs* [9], nos quais uma face expressiva inicial é obtida a partir da deformação dos componentes faciais, e as outras alterações nas imagens são sintetizadas usando *GANs*. Outro exemplo de uma abordagem híbrida é a *Deferred Neural Rendering* [5]. Nesta abordagem, os vetores de características dos mapas de textura, chamados de texturas neurais, são aprendidos usando um modelo semelhante ao U-Net. Em seguida, os mapas aprendidos são combinados com um pipeline gráfico tradicional por meio de deformações.

As abordagens baseadas em Aprendizado de Máquina apresentam resultados limitados quando a expressão facial desejada possui poucos exemplos para treinamento [1]. Já as algumas abordagens híbridas e baseadas em Deformação e Mapeamento utilizam apenas as imagens de um indivíduo, escolhidas de forma manual, como fonte para gerar a nova imagem, porém essa abordagem fica bastante dependente da imagem fonte escolhida. Uma escolha inadequada pode gerar inconsistências nos resultados [17], [18], o que pode ser especialmente problemático para expressões faciais menos comuns [17], [18] e espontâneas [6]. Uma maneira de contornar essa limitação nessas abordagens é selecionar uma imagem fonte semelhante a D_n, o que pode reduzir inconsistências indesejadas, mantendo as características da face de destino e usando apenas uma amostra como imagem fonte. Poucos trabalhos se preocuparam em selecionar essas imagens para otimizar seus resultados [1]. Alguns se preocupam apenas com as regiões oculares [17], [18], enquanto o presente estudo considera toda a expressão facial. Outros artigos consideram a disponibilidade de um conjunto de imagens com variadas expressões faciais da face destino [19], [20], o que pode restringir as aplicações das imagens sintéticas.

Referências

[1] R. L. Testa, C. G. Corrêa, A. Machado-Lima, and F. L. S. Nunes, "Synthesis of facial expressions in photographs: Characteristics, approaches, and challenges," ACM Comput. Surv., vol. 51, no. 6, pp. 124:1–124:35, Jan. 2019. [Online]. Available: http://doi.acm.org/10.1145/3292652

[2] D. Mima, H. Kubo, A. Maejima, and S. Morishima, "Automatic generation of facial wrinkles according to expression changes," in SIGGRAPH Asia 2011 Posters, ser. SA '11. New York, NY, USA: ACM, 2011, pp. 1:1–1:1.

[3] W. Wei, C. Tian, S. J. Maybank, and Y. Zhang, "Facial expression transfer method based on frequency analysis," Pattern Recognition, vol. 49, pp. 115 – 128, 2016.

[4] J. Ghent and J. McDonald, "Photo-realistic facial expression synthesis," Image and Vision Computing, vol. 23, no. 12, pp. 1041 – 1050, 2005.

[5] J. Thies, M. Zollhöfer, and M. Nießner, "Deferred neural rendering: Image synthesis using neural textures," ACM Trans. Graph., vol. 38, no. 4, pp. 66:1–66:12, Jul. 2019. [Online]. Available: http://doi.acm.org/10.1145/3306346.3323035

[6] Y. Zhou and B. E. Shi, "Photorealistic facial expression synthesis by the conditional difference adversarial autoencoder," in 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), Oct 2017, pp. 370–376.

[7] Y. Choi, M. Choi, M. Kim, J. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," CoRR, vol. abs/1711.09020, 2017. [Online]. Available: http://arxiv.org/abs/1711.09020

[8] L. Song, Z. Lu, R. He, Z. Sun, and T. Tan, "Geometry guided adversarial facial expression synthesis," in Proceedings of the 26th ACM International Conference on Multimedia, ser. MM '18. New York, NY, USA: ACM, 2018, pp. 627–635. [Online]. Available: http://doi.acm.org/10.1145/3240508.3240612

[9] J. Geng, T. Shao, Y. Zheng, Y. Weng, and K. Zhou, "Warp-guided gans for single-photo facial animation," ACM Trans. Graph., vol. 37, no. 6, pp. 231:1–231:12, Dec. 2018. [Online]. Available: http://doi.acm.org/10.1145/3272127.3275043

[10] H. Ding, K. Sricharan, and R. Chellappa, "Exprgan: Facial expression editing with controllable expression intensity," in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[11] D. Marčetić, M. Soldić, and S. Ribarić, "Hybrid cascade model for face detection in the wild based on normalized pixel difference and a deep convolutional neural network," in Computer Analysis of Images and Patterns, M. Felsberg, A. Heyden, and N. Krüger, Eds. Cham: Springer International Publishing, 2017, pp. 379–390.

[12] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in 2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Oct 2018, pp. 471–478.

[13] N. Wang, X. Gao, D. Tao, H. Yang, and X. Li, "Facial feature point detection: A comprehensive survey," Neurocomputing, pp. –, 2017.

[14] H. Averbuch-Elor, D. Cohen-Or, J. Kopf, and M. F. Cohen, "Bringing portraits to life," ACM Trans. Graph., vol. 36, no. 6, pp. 196:1–196:13, Nov. 2017. [Online]. Available: http://doi.acm.org/10.1145/3130800.3130818

[15] R. L. Testa, A. Machado-Lima, Nunes, and F. L. S., "Factors influencing the perception of realism in synthetic facial expression," in 2018 31th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Oct 2018, pp. 127–133, (in press).

[16] Z. Liu, Y. Shan, and Z. Zhang, "Expressive expression mapping with ratio images," in Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, ser. SIGGRAPH '01. New York, NY, USA: ACM, 2001, pp. 271–276.

[17] L. Xiong, N. Zheng, Q. You, J. Liu, and S. Du, "Eye synthesis using the eye curve model," in 19th IEEE International Conference on Tools with Artificial Intelligence(ICTAI 2007), vol. 2, Oct 2007, pp. 531–534.

[18] L. Xiong, N. Zheng, J. Liu, S. Du, and Y. Liu, "Eye synthesis using the eye curve model," Image and Vision Computing, vol. 28, no. 3, pp. 329 – 342, 2010.

[19] K. Li, F. Xu, J. Wang, Q. Dai, and Y. Liu, "A data-driven approach for facial expression synthesis in video," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 2012, pp. 57–64.

[20] K. Li, Q. Dai, R. Wang, Y. Liu, F. Xu, and J. Wang, "A data-driven approach for facial expression retargeting in video," IEEE Transactions on Multimedia, vol. 16, no. 2, pp. 299–310, Feb 2014.