

Sistema de Reconhecimento de Expressões Faciais

A. L. Zanellato*, C. B. Ferreira*, H. B. Gamba**, G. B. Borba*, V. Pilla Jr*

*Departamento de Eletrônica - **Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial, Universidade Tecnológica Federal do Paraná, Curitiba, Brazil
e-mail: vpillajr@gmail.com

Introdução: As expressões faciais constituem uma importante forma de comunicação humana não verbal. Sistemas de reconhecimento de expressões faciais são aplicáveis na detecção de fadiga, na identificação suspeitos, em ambientes de aprendizado e de entretenimento interativos, entre outras aplicações. Neste contexto, o objetivo deste projeto é classificar a expressão facial presente em uma imagem frontal da face humana entre as sete expressões humanas reconhecíveis exclusivamente pela imagem da face: felicidade, tristeza, medo, raiva, surpresa, aversão e neutro [1]. A presente abordagem fez uso de uma rede neural profunda constituída por uma estrutura neural convolucional e uma segunda estrutura neural como classificador, inspirada em [2]. A arquitetura foi implementada com o ambiente *Caffe* (<http://http://caffe.berkeleyvision.org/>). O processo de treinamento e validação fez uso do *conjunto de dados* Cohn-Kanade estendido (CK+) (<http://www.pitt.edu/~emotion/ck-spread.htm>). A partir da arquitetura treinada, desenvolve-se um sistema com operação em tempo real.

Métodos: O conjunto de dados original CK+ contém sequências de imagens entre a transição da expressão neutra para a expressão de uma emoção específica. Foram utilizadas 3.406 imagens de 123 pessoas, distribuídas de forma desbalanceada entre as classes. Às imagens, transformadas para escala de cinza (8 bits por pixel), aplicou-se um recorte de 256×256 pixels centralizados pela face com o auxílio do algoritmo Viola-Jones. O conjunto de dados foi submetido a um processo de *data augmentation*, realizado com as seguintes transformações: (a) compressão JPG (qualidades de 20%, 15% e 10%); (b) equalização de histograma; (c) adição de ruído Gaussiano com média 0 e variância 0,01; (d) níveis L_1 e L_2 da decomposição multiescala por meio de pirâmide Gaussiana; (e) *grayscaleing* com limiares de 128, 64, 32 e 16 níveis; (f) espelhamento. Combinações de mais de uma transformação também foram aplicadas à cada imagem, resultando em um conjunto de dados final de 149.864 imagens, das quais se constituíram os conjuntos de treinamento e de validação. A arquitetura de rede neural profunda dispõe de cinco camadas convolucionais, duas camadas totalmente conectadas e uma camada de saída *softmax*, descrita detalhadamente em [2]. O treinamento emprega a validação cruzada *k-fold*, com $k = 10$. Desenvolveu-se um sistema em tempo real que classifica expressões faciais em *streaming* de vídeo, composto por cinco redes treinadas e um sistema de votação para seleção do rótulo.

Resultados: O treinamento empregando a validação cruzada *10-fold* resulta em 10 redes neurais. A acurácia média obtida para as classes são: (a) neutro, 95,65%; (b) raiva, 93,18%; (c) aversão, 98,80%; (d) medo, 89,87%; (e) felicidade, 98,42%; (f) tristeza, 90,74% e (g) surpresa, 97,57%. Além da acurácia, a sensibilidade e o valor preditivo positivo também foram considerados na avaliação de todas as 10 redes para cada emoção, sendo possível observar que algumas redes neurais, apesar de alcançarem acurácias altas, são significativamente inferiores às demais na classificação de determinadas expressões. O sistema em tempo real implementado mostrou-se funcional. No sistema em tempo real as expressões faciais são analisadas e rotuladas quadro a quadro, a partir do *streaming* de vídeo.

Conclusão: As redes neurais profundas desenvolvidas neste trabalho alcançaram a acurácia média de 92,24% na classificação de expressões faciais do conjunto de validação. O sistema para classificação de expressões faciais em tempo real implementado até o momento é funcional.

Referências: [1] Keltner D and Ekman P, Handbook of Emotions, ch. 15 - Facial Expression of Emotion, pp. 151–249. Guilford Publications, Inc., 2nd ed., 2000. [2] Krizhevsky A, Sutskever I, and Hinton GE, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25 (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.