

Desenvolvimento Preliminar de um Algoritmo de Detecção em Tempo Real de Estados Emocionais através de Redes Neurais Convolucionais

Gabriel A. R. Silva, Marco A. R. Alves, Bruno C. Bispo e Pedro M. Rodrigues

Resumo— Este trabalho apresenta os resultados preliminares de um algoritmo em desenvolvimento para detecção em tempo real de estados emocionais através da análise de imagens do rosto. Para avaliar o estado emocional, o algoritmo de Viola-Jones é aplicado para segmentar o rosto e uma rede neural convolucional (CNN) é utilizada para classificar as imagens analisadas. Os resultados demonstram uma acurácia de 96,18% na classificação do grupo de teste. Ainda assim, o algoritmo será otimizado para uma melhor detecção e a base de dados utilizada deverá ser ampliada de forma a generalizar os resultados.

Palavras-Chave— Detecção de estados emocionais, algoritmo de Viola-Jones, rede neural convolucional.

Abstract— This paper presents the preliminary results of an under development algorithm for emotional states real-time detection through an analysis of face images. In order to evaluate the emotional state, the Viola-Jones algorithm is applied to segment the face and a convolutional neural network (CNN) is used to classify the analyzed images. The results show an accuracy of 96.18% for the classification of the test group. The algorithm still need to be optimized for better face detection and the used database must be extended in order to generalize the results.

Keywords— Emotion detection, Viola-Jones algorithm, convolutional neural network.

I. INTRODUÇÃO

As emoções são estados mentais e psicológicos que envolvem pensamentos e sentimentos que influenciam os comportamentos e ações de indivíduos. A análise de estados emocionais pode ser utilizada para, por exemplo, a monitorização e vigilância, em robôs inteligentes, nas interfaces computador-homem e em sistemas de alerta ao condutor [1]. Este trabalho apresenta os resultados preliminares do desenvolvimento de um algoritmo para avaliar o estado emocional de um indivíduo em tempo real, através da sua análise facial recorrendo a uma rede neural convolucional (CNN).

II. METODOLOGIA

A metodologia desenvolvida envolve três etapas: aquisição digital de imagem e pré-processamento; segmentação do rosto e extração de características; Análise em tempo real. Na Figura 1 está apresentada toda a metodologia aplicada.

Gabriel A. R. Silva, Marco A. R. Alves, e Pedro M. Rodrigues, Universidade Católica Portuguesa, CBQF - Centro de Biotecnologia e Química Fina - Laboratório Associado, Escola Superior de Biotecnologia, Porto, Portugal, e-mail: pmrodrigues@ucp.pt; Bruno C. Bispo, Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil, e-mail: bruno.bispo@ufsc.br.

A. Aquisição e pré-processamento de imagem

Nesta primeira etapa foi utilizada uma *webcam* para adquirir 6929 imagens do rosto de 10 indivíduos que participaram neste estudo. Para cada indivíduo, foram adquiridas imagens onde os mesmos expressavam três tipos de emoções nomeadamente felicidade, tristeza e neutralidade. As imagens, na escala RGB, foram filtradas e convertidas para a escala de cinzas.

B. Segmentação do rosto e extração de características

A identificação e segmentação do rosto foi feita através do algoritmo de detecção de objetos de Viola-Jones. Este algoritmo tem a vantagem de possibilitar um processamento rápido e preciso na detecção de estruturas do rosto, sendo muito utilizado em análises em tempo real [2], [3]. O algoritmo Viola-Jones é dividido em: características Haar; imagens integrais; treinamento *AdaBoost* e classificadores em cascata [2], [3].

1) *Caraterísticas Haar*: As caraterísticas Haar são baseadas em funções Haar e permitem identificar regiões com diferentes intensidades no rosto, estando estas associadas a estruturas mais claras ou mais escuras. Neste processo são utilizados 3 tipos de características, compostas por matrizes binárias com regiões adjacentes vertical ou horizontalmente de tamanho idêntico que permitem detectar estruturas como limites, linhas horizontais, verticais e oblíquas. Ao aplicar estas características a uma janela da imagem, a intensidade dos pixels das regiões claras e escuras é somada e a diferença entre regiões é calculada. Valores elevados deste processo indicam uma semelhança entre a característica e a janela da imagem analisada.

2) *Imagens integrais*: A imagem integral é uma representação transitória da imagem original utilizada para tornar a extração de características Haar mais rápida e eficiente. O valor de cada pixel nesta representação resulta da soma do seu próprio valor com os pixels acima e à sua esquerda.

3) *Treinamento AdaBoost*: Nesta etapa um algoritmo *Adaptive Boosting* é utilizado para selecionar o melhor subconjunto de características. Cada característica representa um classificador fraco, sendo verificada a sua performance em todas as subregiões da imagem. O peso de cada classificador é ajustado de acordo com a sua performance e um classificador forte ou *boosted classifier* é criado, combinando os classificadores fracos com menos erro associado. Este processo torna o algoritmo mais preciso já que permite detectar falsos positivos e verdadeiros negativos presentes nos dados.

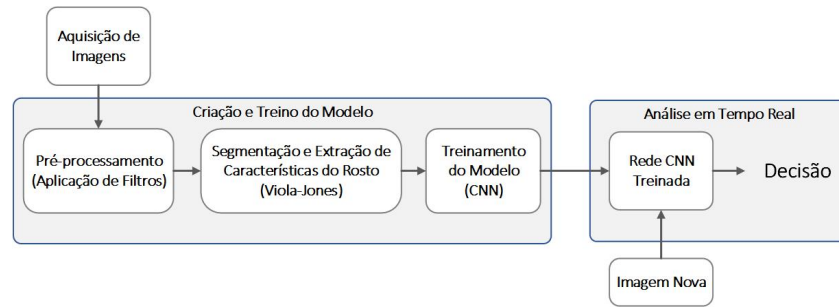


Fig. 1. Diagrama de fluxo da metodologia aplicada.

4) *Classificadores em cascata*: Estes classificadores são utilizados para melhorar a velocidade e precisão do modelo. Cada subjanela das subregiões entra num processo iterativo no qual se verifica a presença das características fracas do classificador forte por ordem de importância. A ausência de qualquer uma destas características numa subjanela, leva a que se descarte a presença da estrutura que se pretende detectar na respetiva subregião da imagem. Desta forma é possível descartar rapidamente as regiões da imagem onde não se encontram estruturas de interesse.

C. Análise em tempo real

De forma a realizar uma detecção do estado emocional do indivíduo em tempo real, foi utilizada uma CNN, sendo estes classificadores reconhecidos pela sua capacidade de analisar e identificar padrões significativos em imagens [4]. Neste estudo recorreu-se a uma CNN baseada na arquitetura da rede *AlexNet*, tendo-se classificado as expressões faciais de felicidade, tristeza e neutralidade. A CNN utilizada é composta por 25 camadas que incluem *Convolutional layers*, *Maxpooling layers*, *Fully Connected layers* e *Dropout layers*, com funções de activação ReLU (*rectified linear unit*) excepto na última camada onde foi utilizada uma função *softmax*. Um total de 80% das 6929 imagens foram utilizadas como par entrada/saída para treinar a CNN, tendo-se utilizado o algoritmo RMSProp (*root mean square propagation*) e uma taxa de aprendizagem de 0,00001 para esse efeito.

III. RESULTADOS E DISCUSSÃO

Nesta fase preliminar, o treinamento do modelo ajustou-se bem aos dados, permitindo atingir um valor de acurácia de 96.18% na classificação do grupo de teste. O algoritmo foi testado em tempo real em indivíduos que não participaram do processo de treinamento e teste, estando apresentado na Figura 2 exemplos do mesmo. Em cada imagem o rosto detectado é identificado por uma janela retangular que mostra também o resultado da classificação do sistema. Contudo, há a necessidade de uma boa luminosidade e estabilidade das estruturas de interesse na imagem para uma correta detecção, sendo necessário otimizar futuramente a detecção do rosto.

IV. CONCLUSÕES

O presente trabalho apresenta os resultados preliminares do desenvolvimento de um algoritmo para detecção do estado



Fig. 2. Exemplos do algoritmo funcionando em tempo real.

emocional de um indivíduo através da análise de imagens do rosto. O algoritmo de Viola-Jones foi aplicado para identificar e segmentar imagens do rosto que foram utilizadas para treinar uma CNN - baseada na arquitetura da rede *AlexNet* e para realizar uma detecção em tempo real. Apesar dos bons resultados preliminares, é necessário aplicar o algoritmo desenvolvido a uma base de dados mais extensa de forma generalizar os resultados. Finalmente, é pretende-se futuramente otimizar o algoritmo de forma a diminuir erros por alterações na luminosidade e estabilidade das imagens.

AGRADECIMENTOS

Esse trabalho foi apoiado pelos Fundos Nacionais da Fundação para a Ciência e a Tecnologia (FCT), Portugal, através do projeto UIDB/50016/2020.

REFERÊNCIAS

- [1] D. Reney and N. Tripathi, "An efficient method to face and emotion detection," in *2015 Fifth International Conference on Communication Systems and Network Technologies*. IEEE, apr 2015.
- [2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR 2001. IEEE Comput. Soc, feb 2001.
- [3] M. N. Chaudhari, M. Deshmukh, G. Ramrakhiani, and R. Parvatkar, "Face detection using viola jones algorithm and neural networks," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA)*. IEEE, aug 2018.
- [4] R. Chauhan, K. K. Ghanshala, and R. Joshi, "Convolutional neural network (CNN) for image detection and recognition," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. IEEE, dec 2018.