

Deteção de Sonolência em Condutores

Gonçalo Santos, MEEC, FEUP, up202003537@up.pt

Guilherme Moreira, MEEC, FEUP, up201806631@up.pt

Pedro Ribeiro, MEEC, FEUP, up201806749@up.pt

Resumo—Este relatório apresenta o desenvolvimento e a implementação de um algoritmo de deteção de sonolência em condutores. Este relatório irá abordar as metodologias escolhidas, nomeadamente a segmentação da cor da pele, a deteção das faces e bocas, tal como a análise dos resultados obtidos.

I. INTRODUÇÃO

No âmbito da unidade curricular de Visão por Computador, foi-nos proposto o desenvolvimento de um projeto que conseguisse detetar o cansaço de um condutor com base se este estivesse a bocejar ou não. Este projeto encontra-se dividido em três tarefas.

A primeira tarefa consiste em detetar a cara do condutor e criar uma *bounding box* para obter uma máscara. A seguir, a segunda tarefa consiste em detetar, com base no *ground truth* das caras dos condutores, as bocas e classificar as imagens segundo “yawn” ou “no yawn”. Por último, a terceira parte consiste em repetir a tarefa 2, mas para as máscaras geradas na primeira parte.

Para o desenvolvimento deste algoritmo foram usadas operações morfológicas e segmentação de imagens, tal como o método de *Jaccard* para avaliar os resultados obtidos na primeira tarefa e a criação de uma matriz de confusão para avaliar os resultados da segunda e terceira parte.

II. METODOLOGIA

A. Segmentação da cor da pele

Para a deteção da cara do condutor foi optado por segmentar a imagem consoante a cor da pele. Tendo isto em conta, as imagens foram convertidas para os espaços de cores RGB, que consiste na junção da componente vermelha, verde e azul da imagem, e YCbCr, que consiste na componente da luminância e na diferença da informação cromática da cor azul e vermelha em referência à luminância da imagem.

Um dos fatores considerados na nossa solução foi a condição de luminosidade da imagem. A pesquisa em artigos que abordavam este tema revelou que diferentes condições de luz ambiente produzem tonalidades diferentes, afetando negativamente a posterior deteção da pele. No sentido de mitigar este efeito e fazendo recurso da componente Y do espaço de cores YCbCr, pixéis com valores baixos de luminância foram compensados. [1]

No algoritmo desenvolvido, cada uma das componentes de RGB e YCbCr da imagem original são analisadas segundo um

grupo de regras, apresentadas em (1) e (2), obtidas a partir da pesquisa em artigos e estudos efetuados sobre a segmentação da cor da pele. [1][2]

$$(1) (R > 95) \& (G > 40) \& (B > 20) \\ \& ((\max(R, G, B) - \min(R, G, B)) > 15) \\ \& (|R - G| > 15) \\ \& (R > G) \& (R > B)$$

$$(2) 60 < Y < 255 \& \\ 100 < Cb < 125 \& \\ 135 < Cr < 170$$

Caso estas condições se verifiquem, esse pixel é adicionado à máscara da imagem, como é apresentado na Figura 2, que resulta da imagem original apresentada na Figura 1.



Figura 1 - Imagem Original

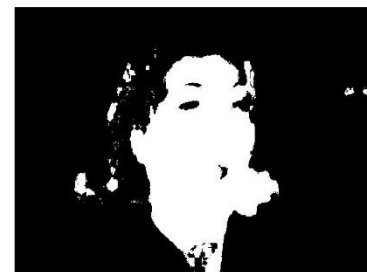


Figura 2 - Resultado da Segmentação da Pele

B. Deteção e segmentação da face

De modo a atenuar algum “ruído” residual criado pelo algoritmo de deteção de cor e remover falhas no interior da cara foram aplicadas operações morfológicas à máscara, como recomendado na pesquisa efetuada. Desta máscara é retirada a região de maior dimensão, que se assume ser a região que representa a cara (Figura 3). [1][3]



Figura 3 - Região com maior dimensão detetada

Com base nas propriedades desta região, em particular o centroide e os tamanhos do eixo maior e do eixo menor da elipse que melhor se aproxima à região, definem-se as dimensões precoces da *bounding box*, assumindo que a altura e a largura desta têm, respetivamente, o mesmo comprimento do eixo maior e do eixo menor.

Com o intuito de corrigir defeitos no processo de segmentação que não tenham sido corrigidos nas operações morfológicas, como deteção de zonas de pele que não fazem parte da cara ou a eventual deteção de cabelo e peças de roupa, as dimensões da *bounding box* são ajustadas para melhor envolver a zona da cara. Nos casos em que mais zonas são consideradas, estas tendem a estar posicionadas abaixo da zona da cara. Como tal, a posição vertical do canto superior esquerdo da *bounding box* é fixada e a altura é alterada para 80% da sua dimensão se esta for superior a 75% da altura da imagem, ou se a posição vertical do lado inferior da *bounding box* estiver abaixo de 88% da altura da imagem. Em relação à largura, é considerado que uma *bounding box* com largura superior a 40% da largura da imagem será demasiado grande e é, por isso, ajustada para 60% do tamanho originalmente definido.

Por fim, define-se a *bounding box* final, usada para criar a máscara (Figura 4).



Figura 4 - Bouding Box da Cara

C. Deteção e segmentação da boca

Para a deteção da boca baseámo-nos na pesquisa feita de que a cara na zona da boca contem vermelhos mais intensos que os azuis, ou seja, que a componente C_r é superior à componente C_b . Deste modo, aplicamos as fórmulas apresentadas em (3) e (4), onde C_r^2 e C_r/C_b são normalizados para o intervalo [0, 255] e n representa o tamanho da máscara. [4][5]

$$(3) Mouth_{map} = C_r^2 \cdot (C_r^2 - \eta \cdot \frac{C_r}{C_b})^2$$

$$(4) \eta = \frac{\frac{1}{n} \sum_{(x,y) \in FG} C_r(x,y)^2}{\frac{1}{n} \sum_{(x,y) \in FG} \frac{C_r(x,y)}{C_b(x,y)}}$$

A seguir, é possível verificar o resultado a demonstração de C_r^2 (Figura 5) e C_r/C_b (Figura 6) tal como da fórmula em (3) (Figura 7).



Figura 5 - C_r^2



Figura 6 - C_r/C_b



Figura 7 - Mouthmap

Para binarizar o resultado obtido na Figura 7, com o intuito de retirar apenas a boca, é usado inicialmente um *threshold* obtido a partir da intensidade média dos pixels. Se com este *threshold* não for possível obter qualquer região a branco, o *threshold* é reduzido sucessivamente para 80% do seu valor até a obter.

Após a deteção da boca procede-se à sua segmentação. Para tal, na imagem RGB, com as dimensões da máscara previamente obtida, é recortada, em torno da boca, uma zona retangular com 70% da largura e 30% da altura da máscara e limitada pela mesma (Figura 8).



Figura 8 - Resultado da Segmentação da Boca

D. Detecção de “yawn” e “no yawn”

Para classificar as imagens segundo “yawn” ou “no yawn”, considerou-se, com base na literatura encontrada, que quando uma pessoa boceja, o interior da sua boca fica mais escura do que o resto da face, isto é, no espaço de cores RGB, o valor de $R - G$ é inferior aos lábios e à pele (Figura 9). [6]

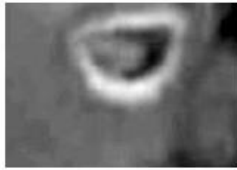


Figura 9 - R-G

Deste modo, o interior da boca (Figura 10) é obtido através da fórmula em (5) que recorre a um *threshold* (t) adaptativo, calculado da seguinte forma:

1. Cálculo de $R - G$ na zona da boca;
2. Criação de uma máscara com os 10% pixéis de valores mais elevados de $R - G$;
3. Remoção da zona de maior dimensão desta nova máscara (boca), isto é, obtenção da pele nos 10% pixéis obtidos no passo anterior;
4. Caso este número de pixéis de pele seja maior que 5% do tamanho da zona da boca, definir o *threshold* como sendo o valor máximo da pele calculada no passo 3, caso contrário passar ao passo seguinte;
5. Junção dos 1/15 pixéis de valores mais inferiores de $R - G$ com os pixéis obtidos no passo 3 (pele) e voltar ao passo 2 até que se verifique a condição do passo 4;

$$(5) h_{(x,y)} = \begin{cases} 1, & R - G \leq t \\ 0, & R - G > t \end{cases}$$



Figura 10 - Resultado da Segmentação do Interior da Boca

Após a aplicação destas operações, idealmente obtém-se uma imagem binária somente com o interior da boca, mas o que realmente acontece é que é obtido o interior da boca com algum ruído resultante de outras partes escuras. (Figura 10)

De modo a corrigir este problema, retira-se da imagem as zonas de maior dimensão e verifica-se se alguma destas zonas tem o centroide aproximadamente no centro da imagem. Caso isso se verifique, classificamos a imagem como sendo “yawn” (Figura 11).



Figura 11 - Resultado Final da Segmentação do Interior da Boca

III. DISCUSSÃO E ANÁLISE DE RESULTADOS

A. Avaliação das máscaras geradas

Tabela 1 - Métricas do Training Dataset Task 1

Training Task 1	
Recall	0,8750
Precision	0,8750
F1-Measure	0,8750
TP	35
FP	5
FN	5
Total	40

Tabela 2 - Métricas do Test Dataset Task 1

Test Task 1	
Recall	0,8500
Precision	0,8500
F1-Measure	0,8500
TP	17
FP	3
FN	3
Total	20

Os resultados obtidos na primeira parte do projeto (detecção e segmentação da cor da pele) foram satisfatórios em ambos *datasets*, como observado nas tabelas 1 e 2. Apesar das limitações no processo de detecção da cara pela tonalidade da pele, como a detecção de partes com cores semelhantes, as operações morfológicas e as operações de ajuste das dimensões da *bounding box* posteriores foram eficazes, na maioria dos casos, a conseguir gerar uma *bounding box* adequada à imagem.

Tal como referido anteriormente, foi aplicado o índice de Jaccard para avaliar a máscara gerada da *bounding box* proveniente da detenção e segmentação da face com a máscaras fornecidas, que serviram como o *ground truth*.

B. Avaliação da classificação “yawn” vs “no yawn”

Tabela 3 - Matriz de Confusão e Métricas do Training Dataset Task 2

Training Task 2		Ground Truth		Total
		Yawn	No Yawn	
Algoritmo	Yawn	19	10	29
	No Yawn	1	10	11
	Total	20	20	40

Recall	0,9500
Precision	0,6552
F1-Measure	0,7755

Tabela 4 - Matriz de Confusão e Métricas do Test Dataset Task 2

Test Task 2		Ground Truth		Total
		Yawn	No Yawn	
Algoritmo	Yawn	6	4	10
	No Yawn	4	6	10
	Total	10	10	20

Recall	0,6000
Precision	0,6000
F1-Measure	0,6000

Na segunda tarefa do projeto, a classificação e segmentação da boca, com base no *dataset* fornecido de máscaras, geraram resultados mistos, como apresentado nas tabelas 3 e 4.

Na detecção e segmentação da boca, a ideia que a cara na zona da boca contem vermelhos mais intensos provou estar correta, mas não tem em conta que as máscaras não são perfeitas e contêm outras partes da imagem que podem ser tão ou mais vermelhas que os lábios, por exemplo, roupa vermelha. Nestes casos, após criar a imagem binária, caso haja alguma zona de área superior à dos lábios será essa a selecionada, incorretamente, como zona da boca.

Por um lado, se o interior da boca da imagem fornecida possuir tons escuros, o algoritmo efetua a segmentação e detecção de forma eficaz. Por outro lado, se esta não possuir estes traços, por exemplo, um interior da boca com um nível de luminosidade elevado, o algoritmo não consegue detetar a presença de um “yawn”, classificando incorretamente a imagem como “no yawn”.

C. Avaliação da classificação “yawn” vs “no yawn” com base nas máscaras geradas

Tabela 5 - Matriz de Confusão e Métricas do Training Dataset Task 3

Training Task 3		Ground Truth		Total
		Yawn	No Yawn	
Algoritmo	Yawn	18	8	26
	No Yawn	2	12	14
	Total	20	20	40

Recall	0,9000
Precision	0,6923
F1-Measure	0,7826

Tabela 6 - Matriz de Confusão e Métricas do Test Dataset Task 3

Test Task 3		Ground Truth		Total
		Yawn	No Yawn	
Algoritmo	Yawn	8	7	15
	No Yawn	2	3	5
	Total	10	10	20

Recall	0,8000
Precision	0,5333
F1-Measure	0,6400

Assim sendo, as máscaras geradas pelo algoritmo na primeira tarefa foram postas à prova na detecção de cansaço em ambos os *dataset* fornecidos, tendo revelado métricas interessantes, como apresentadas nas tabelas 5 e 6. As limitações observadas no subcapítulo anterior foram de novo testemunhadas.

IV. CONCLUSÕES

Finalmente, os métodos implementados revelaram cumprir o objetivo inicial proposto e as limitações detetadas poderiam eventualmente ser superadas com a uniformização dos *datasets*, como luminosidade constante e uma posição frontal dos condutores, e/ou um desenvolvimento mais profundo, recorrendo a técnicas de *edge detection* e *feature-based segmentation* ou *image-based segmentation*.

REFERÊNCIAS

- [1] D. Ghimire and J. Lee, “A robust face detection method based on skin color and edges,” *Journal of Information Processing Systems*, vol. 9, no. 1, pp. 141–156, Mar. 2013, doi: 10.3745/JIPS.2013.9.1.141.
- [2] F. Pujol, M. Pujol, A. Jimeno-Morenilla, and M. Pujol, “Face detection based on skin color segmentation using fuzzy entropy,” *Entropy*, vol. 19, no. 1, p. 26, Jan 2017, doi: 10.3390/e19010026.
- [3] A. S. Dhavalikar and R. K. Kulkarni, “Face detection and facial expression recognition system,” 2014 International Conference on Electronics and Communication Systems (ICECS), 2014, doi: 10.1109/ecs.2014.6892834.
- [4] S. Abtahi, B. Hariri, and S. Shirmohammadi, “Driver drowsiness monitoring based on yawning detection,” 2011 IEEE International Instrumentation and Measurement Technology Conference, 2011, doi: 10.1109/imtc.2011.5944101.
- [5] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, “Face detection in color images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, 2002, doi: 10.1109/34.1000242.
- [6] Y. Ji, S. Wang, Y. Lu, J. Wei, and Y. Zhao, “Eye and mouth state detection algorithm based on contour feature extraction,” *Journal of Electronic Imaging*, vol. 27, no. 05, p. 1, 2018, doi: 10.1117/1.jei.27.5.051205.