

SISTEMA PARA DETECÇÃO DE SONOLÊNCIA E BOCEJO EM MOTORISTAS

Felipe M. L. de Moraes¹, Fernando G. M. Veizaga¹, Gabriel S. Silva¹,
Maria I. C. S. Coutinho¹, Vinicius N. Silva¹

¹Departamento de Computação – Centro Federal de Educação
Tecnológica de Minas Gerais (CEFET-MG)
30510-000 – Belo Horizonte – MG – Brasil

Abstract.

Resumo. *Este artigo aborda a problemática da sonolência no trânsito, um tema de grande relevância para a segurança nas vias. A sonolência, um estado inconsciente de sono, tem sido identificada como um fator significativo nos acidentes automobilísticos, resultando em um sério risco para condutores e demais usuários das vias. Neste artigo, propomos uma abordagem tecnológica para enfrentar esse problema.*

1. Introdução

Os termos sonolência e bocejos correspondem ao estado de sono, que por sua vez é um estado inconsciente de um ser humano. Dito isso, diversas pesquisas atestam a influência desses termos nos acidentes de trânsito. De acordo com [WHO 2022], ocorrem 1,3 milhões de acidentes de trânsito por ano, e a sonolência é apontada como a principal causa de mortes nesses acidentes [Hemadri and Kulkarni 2013]. Assim, a sonolência no trânsito é uma questão de extrema preocupação no cenário contemporâneo. A falta de atenção, concentração e tempo de reação devido à sonolência pode ter consequências catastróficas nas vias públicas.

Uma importante abordagem para solucionar problemas relacionados a esse tema é a utilização da visão computacional. A visão computacional é um campo da inteligência artificial que se concentra no processamento de informações visuais [Poggio et al. 1985]. Esse ramo tem avançado rapidamente nas últimas décadas, impulsionado pelo aumento do poder computacional, pela disponibilidade de grandes conjuntos de dados e pelos avanços em algoritmos e técnicas de aprendizado de máquina. Esses sistemas são capazes de realizar uma variedade de tarefas, como detecção e reconhecimento de objetos, segmentação de imagens, reconhecimento de padrões, bem como análise de expressões.

Nesse contexto, o presente artigo propõe o uso de técnicas de inteligência artificial para a identificação de sonolência e bocejos, visando identificar sinais de sonolência e alertar o condutor em tempo hábil. O artigo está organizado da seguinte forma: na segunda seção, são apresentadas as especificações do trabalho, incluindo descrição das bibliotecas e a análise de alguns pontos críticos; na terceira seção, é explicitado a metodologia utilizada na implementação do programa; na quarta seção, os resultados obtidos com o experimento e, por fim, o artigo conclui com considerações relevantes sobre o estudo realizado, destacando as contribuições, limitações e possíveis direções futuras para a pesquisa.

2. Especificação dos processos

Para a realização dos procedimentos, utilizou-se a linguagem Python, reconhecida por sua popularidade em Machine Learning e Inteligência Artificial, com um foco principal nas bibliotecas OpenCV e dlib.

O OpenCV é uma biblioteca de visão computacional e aprendizado de máquina de código aberto que foi construída para fornecer uma infraestrutura comum para aplicações de visões computacionais. Esta biblioteca tem uma grande quantidade de algoritmos de otimização, os quais são capazes de detectar e reconhecer faces, identificar objetos, classificar ações humanas em vídeos e até extrair modelos 3D de objetos. Já o Dlib é uma ferramenta moderna produzida em C++ que possui algoritmos e ferramentas de aprendizado de máquina para resolver problemas do mundo real.

No contexto desta aplicação, o OpenCV é utilizado para captura de vídeo que possibilita a utilização de inteligência artificial enquanto o dlib foca no reconhecimento facial e nos cálculos para identificação das partes necessárias, como olhos e boca.

3. Metodologia

3.1. Detecção de face

Marcos faciais são usados para localizar e representar regiões importantes da face como olhos, sobrancelhas, nariz, boca e mandíbula. Pontos de referência faciais foram aplicados com sucesso para alinhamento de rosto, estimativa de pose de cabeça, permutação de rosto, detecção de piscadas e muito mais.

A detecção de pontos de referência faciais é um subconjunto do problema de previsão de forma. Dada uma imagem de entrada (e normalmente um ROI (Region of Interest) que especifica o objeto de interesse), um preditor de forma tenta localizar os principais pontos de interesse ao longo da forma.

No contexto dos marcos faciais, o objetivo neste trabalho é detectar estruturas faciais importantes na face usando métodos de previsão de forma. A detecção de pontos de referência faciais é um processo de duas etapas:

- Localizar o rosto na imagem;
- Detectar as principais estruturas faciais na região de interesse do rosto.

A detecção facial pode ser realizada de várias maneiras. No nosso caso, aplicamos um detector de objetos HOG + Linear SVM [TYAGI 2021] pré-treinado especificamente para a tarefa de detecção facial. No entanto, o tipo de algoritmo não é tão importante. O que importa é que, por meio de algum método, obtemos a caixa delimitadora da face (ou seja, as coordenadas (x, y) da face na imagem).

Dada a região da face, podemos então aplicar a Etapa 2: detectar as principais estruturas faciais na região da face.

Há uma variedade de detectores de marcos faciais, mas todos os métodos tentam essencialmente localizar e rotular as seguintes regiões faciais:

- Boca;
- Sobrancelha direita;
- Sobrancelha esquerda;

- Olho direito;
- Olho esquerdo;
- Nariz;
- Mandíbula.

O detector de marco facial incluído na biblioteca dlib é uma implementação do artigo One Millisecond Face Alignment with an Ensemble of Regression Trees de Kazemi e Sullivan [KAZEMI V. 2014].

3.2. Detecção de olhos e lábios

Inicialmente, a região dos olhos se estabelece frequentemente como sendo os seis pontos que compõe o globo ocular, como se vê na Figura 1. No entanto, o uso simples desses pontos apontam a desconsideração da anatomia da pálpebra [DENG and WU 2019].



Figura 1. Pontos críticos da região dos olhos

Para uma melhor aproximação dessa região utiliza-se a técnica de *eye aspect ratio* (EAR) que propõe um maior peso a linha horizontal formada pelos pontos críticos 1 e 4 como mostra a Figura 2. Para que esse maior peso se concretize, torna-se necessário utilizar a equação 1, da qual, a distância ocular horizontal é multiplicada por 2. Ao diminuir-se a altura, a distância horizontal permanece constante de modo que o EAR diminui rapidamente.

$$\frac{||P_2 - P_6|| + ||P_3 - P_5||}{2||P_1 - P_4||} \quad (1)$$

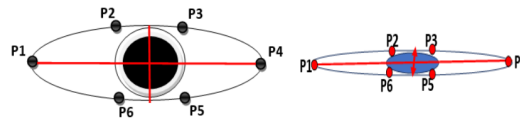


Figura 2. Pontos críticos da região dos olhos

Ademais, é necessário diferenciar o ato de piscar do processo de identificação de sonolência. Conforme a Figura 3, a abertura contínua do olho mostra uma continuidade dos valores de EAR, e quando se há uma piscada, ocorre um vale no histograma, representando o fechamento do olho. Para avaliar a sonolência, o valor do EAR diminui a ponto de não retornar, de modo que o aparelho não reconheça uma abertura de olho e constatando uma sonolência.

Com menos complexidade do que a detecção de olhos, a detecção de lábios ocorre por meio da utilização da distância entre lábios superiores: (i) 49 ao 55 (ii) 61 ao 65, e inferiores: (iii) 56 ao 60 (iv) 66 ao 68, ambos referenciados pela Figura 5. Se torna possível

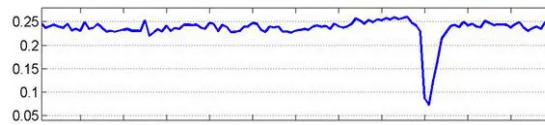


Figura 3. Histograma relativa a continuidade de abertura de olho

realizar esse procedimento pois a cavidade oral não apresenta movimentos mais complicados quando se compara aos olhos. Tanto para a detecção de olhos quanto para lábios existem pesquisas mais sofisticadas como a segmentação binária [DENG and WU 2019].

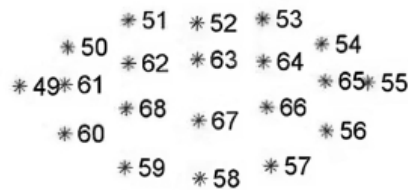


Figura 4. Pontos críticos para os lábios

3.3. Regras de negócio

Regras do negócio, de acordo com [ROSCA D. GREENSPAN S. 1997], são uma nova categoria de requisitos do sistema que englobam decisões sobre como executar o negócio e são caracterizadas pela orientação do negócio e pela sua tendência a mudanças. As regras de negócio ajudam a estabelecer as restrições e validações necessárias para garantir que os dados inseridos e manipulados pelo sistema estejam corretos e em conformidade com as políticas e práticas do negócio. Essa abordagem evita erros, inconsistências e resultados indesejados, proporcionando confiabilidade e precisão às operações do sistema. Diante disso, a aplicação de regras de negócio torna-se indispensável para um funcionamento adequado do sistema.

No contexto específico do sistema em questão, tornou-se necessário estabelecer uma definição para o tamanho mínimo de abertura dos olhos, que serve como um indicativo de que os olhos estão se fechando. Essa definição é crucial e influencia significativamente o sistema, uma vez que as pessoas possuem tamanhos de olhos diversos, variando de pequenos a grandes. Portanto, se o valor atribuído a essa constante for muito alto, o alarme poderá ser acionado desnecessariamente para algumas pessoas, simplesmente por estarem com os olhos abertos.

Além disso, é de extrema importância definir a quantidade consecutiva de detecções em que o olho da pessoa analisada esteja abaixo do tamanho mínimo, indicando que o olho está se fechando. Essa definição é relevante, pois o objetivo do sistema é identificar a sonolência e acionar o alarme antes que a pessoa adormeça, a fim de prevenir acidentes. Se o sistema só alertar após a pessoa já ter adormecido, ele se torna ineficaz em seu propósito.

Da mesma forma, para detectar um bocejo, um dos sintomas de sonolência, é necessário estabelecer uma distância máxima entre os lábios superior e inferior, que serve

como indicativo de que a boca está sendo aberta, juntamente com uma quantidade consecutiva de detecções em que a boca da pessoa analisada esteja acima da distância máxima, indicando a ocorrência de um bocejo. Essas definições são cruciais para que o sistema possa reconhecer e identificar corretamente os bocejos, contribuindo para a detecção precisa de sonolência.

Ademais, outra definição necessária, porém mais óbvia, é estabelecer quando o alarme deve ser ativado e desativado. Enquanto as regras de ativação já foram mencionadas, é igualmente importante definir o momento de desativação desse alarme. Dentre as possibilidades, uma delas seria desativá-lo após um determinado período de tempo, o que aumentaria a probabilidade de a pessoa analisada adormecer, uma vez que o alarme seria desligado em algum momento, tornando-o ineficaz. Outra possibilidade seria exigir o desligamento manual do alarme, mas isso implicaria na integração da câmera com outro sistema físico no veículo, o que pode ser inviável. Por fim, a abordagem utilizada no sistema implementado consiste em desativar o alarme somente quando a pessoa abrir os olhos novamente ou fechar a boca. Essa estratégia visa garantir que o alarme permaneça ativado até que a sonolência seja efetivamente interrompida, evitando acidentes e contribuindo para a segurança dos usuários.

Ao estabelecer os parâmetros corretos para a abertura dos olhos e ocorrência de bocejos, é possível melhorar a precisão e a confiabilidade do sistema, evitando alarmes falsos e assegurando que o sistema seja eficaz na detecção precoce da sonolência. Portanto, a definição adequada das regras de negócio no sistema é fundamental para garantir seu correto funcionamento e sua utilidade na prevenção de acidentes causados pela sonolência no trânsito.

3.4. Otimizações

Em aspectos gerais, a solução apresentada até o momento consegue executar bem em um computador poderoso com 8GB de memória RAM e processador i5 de 8 núcleos. O problema é que esse tipo de detector, em cenário comercial, deveria ser capaz de funcionar bem em hardware modesto como Raspberry Pi. Nesse contexto pode ser necessário tomar algumas medidas para obter ganho de desempenho na execução e, portanto, serão discutidas duas alternativas.

A primeira delas consiste em substituir o algoritmo de detecção de faces HOG + Linear SVM pelo algoritmo Haar Cascade [VIOLA P. 2001]. Esse algoritmo tem como propósito identificar faces baseado em diferenças entre sombras e regiões claras encontradas no rosto de uma pessoa. Por exemplo, a linha dos olhos é mais escura que a região das bochechas e a região dos olhos é mais escura que a região do nariz, ver figura 5. Nesse contexto, são feitas verificações em cascada dessas comparações definidas e caso uma delas não seja atendida, uma grande região é descartada.

Utilizar esse outra abordagem ao invés dos pontos de referência trazem vantagens e desvantagens. Por um lado há ganho de desempenho geral na detecção de face, porém ocorrem mais detecções falsas. Isso acontece porque o algoritmo Haar Cascade é mais sensível a variações de luz, variações do ângulo da câmera e qualidade do vídeo.

A segunda abordagem que pode ser seguida consiste em saltar frames entre verificações. O vídeo em tempo real pode oferecer 60 frames por segundo, a ideia seria fazer a verificação em 30 frames por segundo por exemplo. Nesse contexto são necessários

ajustes dos parâmetros temporais para detecção do eventos.

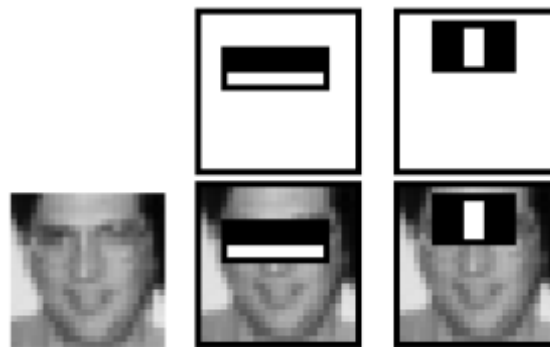


Figura 5. Detecção de face por cascatas

Isso também traz vantagens e desvantagens. Há economia de chamadas de execução para a CPU. Porém isso pode comprometer detecção em tempo real e pode dificultar o ajuste. Vale lembrar que os frames são referência acusar eventos de sonolência e bocejo, dessa forma pequenas variações nos parâmetros de frames consecutivos provocam grandes variações no tempo de detecção.

4. Resultados

Uma vez configuradas as condições para o início de testes, foram observadas algumas interessantes conclusões acerca do algoritmo em exposição.

Após a análise dos resultados obtidos pelo algoritmo de detecção de sonolência nos olhos e bocejos, utilizando as dimensões dos olhos e a abertura da boca como parâmetros, podemos chegar a resultados interessantes.



Figura 6. Corretamente identificando olhos fechados

Inicialmente, quando as condições estavam configuradas para um estado considerado normal, ou seja, olhos bem abertos, boca fechada e todos os elementos faciais dentro do campo de visão da câmera, o procedimento de reconhecimento ocorreu sem problemas, refletindo adequadamente o estado de repouso.

Ao começar a transição do estado de repouso, ou seja, fechando os olhos e/ou abrindo bastante a boca, foi possível observar que os parâmetros configurados para a taxa de frames e o tempo mínimo de reconhecimento de cada "status" atuaram de maneira precisa e efetiva, conforme planejado.

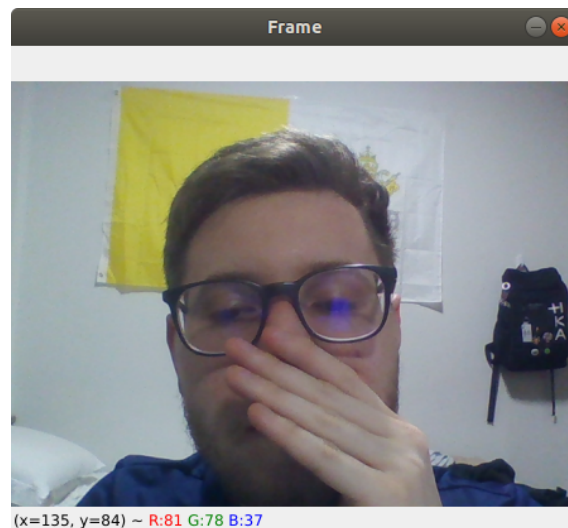


Figura 7. Bocejo utilizando a mão não é identificado pelo algoritmo

Explorando ainda mais as possibilidades do projeto, foi possível ampliar as condições de apresentação do indivíduo nos casos de teste, a fim de avaliar a precisão e a correção do algoritmo desenvolvido.

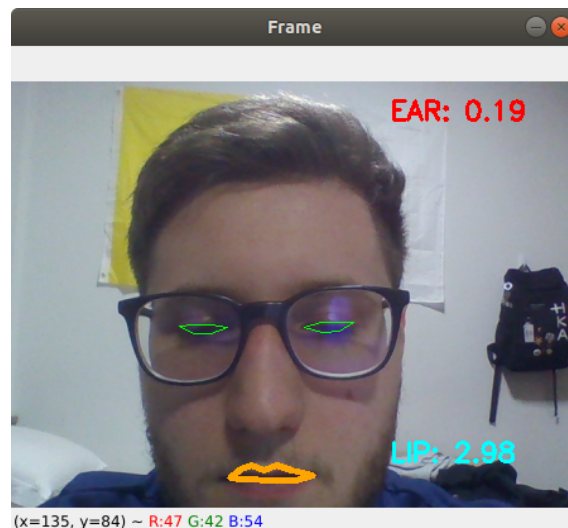


Figura 8. Olhos fechados não sendo identificados

Nesse sentido, é importante destacar que, ao bloquear parcial ou completamente os elementos faciais, o algoritmo não conseguiu reconhecer totalmente as estruturas da boca ou dos olhos, o que naturalmente afetou a detecção de bocejos e/ou sonolência nos olhos. Isso ressalta a necessidade de ter uma visão clara e não obstruída dessas áreas para um funcionamento adequado do algoritmo.

Além disso, não existe uma medida universal segura para a distância entre as pessoas e a câmera, nem uma medida universal para a largura ou altura dos olhos, o que dificulta a identificação da sonolência nos olhos de pessoas com características faciais menores. Isso indica que o algoritmo pode enfrentar desafios ao detectar com precisão os sinais de sonolência em indivíduos com essas características específicas.

Ademais, ações que incluam fechar os olhos continuamente, mesmo que não seja por motivos de sono, também são motivo para acionamento. Assim, cantar ao volante, espirros prolongados ou coceira nos olhos podem ser problemas para o algoritmo orientar corretamente os motoristas.

5. Conclusão

Com os efetivos resultados desta operação balizados pelos resultados obtidos e verificados pelo grupo, o resultado final é bastante satisfatório e pôde ser também comprovado através de demonstração em sala de aula, juntamente com os demais colegas e professor da disciplina.

Portanto, é necessário considerar as limitações supracitadas ao aplicar o algoritmo em diferentes contextos e indivíduos, adaptando-o às condições específicas de cada caso para garantir resultados confiáveis. A pesquisa e o aprimoramento contínuos do algoritmo são fundamentais para superar essas limitações e melhorar sua eficácia em diversas situações.

Referências

- [DENG and WU 2019] DENG, W. and WU, R. (2019). Real-time driver-drowsiness detection system using facial features. *IEEE*, 7.
- [Hemadri and Kulkarni 2013] Hemadri, V. B. and Kulkarni, U. P. (2013). Detection of drowsiness using fusion of yawning and eyelid movements. *Advances In Computing, Communication and Control Communications in Computer and Information Science*, 361:583–594.
- [KAZEMI V. 2014] KAZEMI V., S. J. (2014). One millisecond face alignment with an ensemble of regression trees. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874.
- [Poggio et al. 1985] Poggio, T., Torre, V., and Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317:314–319.
- [ROSCA D. GREENSPAN S. 1997] ROSCA D. GREENSPAN S., FEBLOWITZ M., W. C. (1997). A decision making methodology in support of business rules lifecycle. In *Proceeding of RE 97: IEEE International Symposium on Requirements Engineering*, pages 236–246. IEEE Computer Society Press.
- [TYAGI 2021] TYAGI, M. (2021). Hog (histogram of oriented gradients): An overview.
- [VIOLA P. 2001] VIOLA P., J. M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition*.
- [WHO 2022] WHO, W. H. O. (2022). *Global Status Report on Road Safety*.