

Análise de técnicas de pré-processamento de imagem para reconhecimento facial baseada em VGG Faces e Ball tree

Jeanderson de Sousa Gomes¹, Flavio H.D. Araujo^{1,2}

¹ Departamento de Sistemas de Informação,
Universidade Federal do Piauí - Picos, Brasil

² Programa de Pós Graduação em Engenharia Elétrica,
Universidade Federal do Piauí - Teresina, Brasil

{jeandsgomes10 , flavio86}@ufpi.edu.br

Abstract. *In recent years, facial recognition has become widely present in many devices and systems. However, despite its practicality, the performance of facial recognition is affected by factors such as lighting variation, pose, facial expression and quality of the camera that performs the capture. Therefore, it is necessary to use image pre-processing techniques to deal with these problems. Thus, in this paper, a comparative study of 6 pre-processing methods is performed (grayscale, Gaussian filter, median filter, linear filter, histogram equalization and logarithmic transformation). The tests were performed using the VGG Faces descriptor, and Ball tree as the recognition method. In addition to the bases in their original form, each base was modified by adding artificial noises, which provided new tests that allowed us to investigate the influence of each pre-processing technique in relation to the noise present in the images. The experiments suggest that the use of the median filter produces better results in images with noise such as Salt and Pepper.*

Resumo. *Nos últimos anos o reconhecimento facial se tornou amplamente presente em diversos dispositivos e sistemas. No entanto, apesar de sua praticidade, o desempenho do reconhecimento facial é afetado por fatores como variação de iluminação, pose, expressão facial e qualidade da câmera que realiza a captura. Sendo assim, é necessário o uso de técnicas de pré-processamento de imagens para o tratamento destes problemas. Com isso, neste artigo é realizado um estudo comparativo de 6 métodos de pré-processamento (escala de cinza, filtro gaussiano, filtro da mediana, filtro linear, equalização de histograma e transformação logarítmica). Os testes foram feitos utilizando o descritor VGG Faces, e Ball tree como método de reconhecimento. Além das bases em sua forma original, cada base foi modificada adicionando ruídos artificiais, o que proporcionou novos testes que permitiram investigar a influência de cada técnica de pré-processamento em relação ao ruído presente nas imagens. Os experimentos sugerem que o uso do filtro da mediana produz melhores resultados em imagens com ruídos tipo Sal e Pimenta.*

1. Introdução

O reconhecimento facial possibilita a identificação e distinção de indivíduos através de seus traços faciais, em fotos, vídeos ou em tempo real. Sendo assim, o reconhecimento

facial torna características biométricas em credenciais únicas para o usuário. E por conta da sua praticidade, o reconhecimento facial se tornou presente em diversos dispositivos, como por exemplo, celulares e câmeras digitais, e em diversos setores como: identificação automática de pessoas, na segurança, e na autenticação por reconhecimento facial.

Sua ampla utilização, presente em diversos dispositivos acessíveis a população, facilitaram o surgimento de fatores como tamanho da amostra, variações de poses, iluminação, expressões faciais e qualidade de câmera que realiza a captura. Fatores estes que se tornaram problemas para o desempenho de sistemas de reconhecimento facial. Uma boa abordagem para amenizar, ou até mesmo solucionar esses problemas, é a aplicação de métodos de pré-processamento de imagens. Por meio desses métodos é possível obter imagens faciais em condições satisfatórias para o processo de extração de características, que é uma etapa fundamental para o reconhecimento facial automático [Han et al. 2013].

Com base no exposto, neste artigo apresentamos uma avaliação de diferentes técnicas de pré-processamento de imagens aplicadas ao reconhecimento facial. Nesta avaliação utilizamos 4 bases de imagens, são elas CASIA-WebFace, AT&T, FEI e a UWA Hyperspectral, além disso, também adicionamos ruído sal e pimenta e gaussiano nas imagens das bases. Já as técnicas de pré processamento analisadas foram filtros de escala de cinza, gaussiano, mediana, linear, equalização de histograma e transformação logarítmica. Por fim, para a extração de características das imagens e a etapa de reconhecimento utilizamos a VGG faces e o Ball tree respectivamente.

2. Trabalhos Relacionados

A seguir são descritos alguns trabalhos em aplicações de reconhecimento facial.

[Lee et al. 2020] propõem a utilização do AdaBoost junto a um padrão binário local multibloco (MB-LBP) para a detecção facial, e MB-LBftP juntamente ao Gabor para descrição das características faciais. Esse método alcançou uma acurácia de 97,27% usando o banco de imagens E-face e 99,06% usando o banco de imagens XM2VTS.

[Bah and Ming 2020] propõem um método usando o algoritmo de Padrão Binário Local (LBP) combinado com técnicas de processamento de imagem, como ajuste de contraste, filtro bilateral e equalização de histograma, para solucionar os problemas de variação iluminação, ruído em imagens, escala, pose, etc. O método abordado alcançou uma acurácia de 95%.

[Luo et al. 2021] propuseram um estudo a respeito do uso da Face Augmentation Generative Adversarial Network (FA-GAN). A rede neural proposta tem o intuito de solucionar o problema da má distribuição de dados, presente em bancos de imagens, evitando forte viés na aprendizagem do modelo.

Pela análise dos trabalhos relacionados, percebe-se que existem alguns problemas que afetam a precisão do reconhecimento facial, como variação de iluminação, ruído, escala, nitidez, pose em uma cena, dentre outros. Com isso, o objetivo deste trabalho é contribuir no processo de investigar o uso de técnicas de pré-processamento para o tratamento desses problemas que reduzem o desempenho do reconhecimento facial.

3. Método Proposto

Nesta seção são descritas as etapas que compõem a metodologia proposta neste trabalho. Os estágios descritos são mostrados na Figura 1. A metodologia usada pode ser dividida em 4 etapas: A primeira consiste na aquisição das imagens. Em seguida, é realizado o pré-processamento das imagens. Com as imagens prontas para serem usadas, a próxima etapa é a de extração das características através do descritor VGG Faces. A última etapa consiste no reconhecimento do indivíduo com base nas características extraídas do indivíduo cadastrado, foi utilizado o Ball Tree.

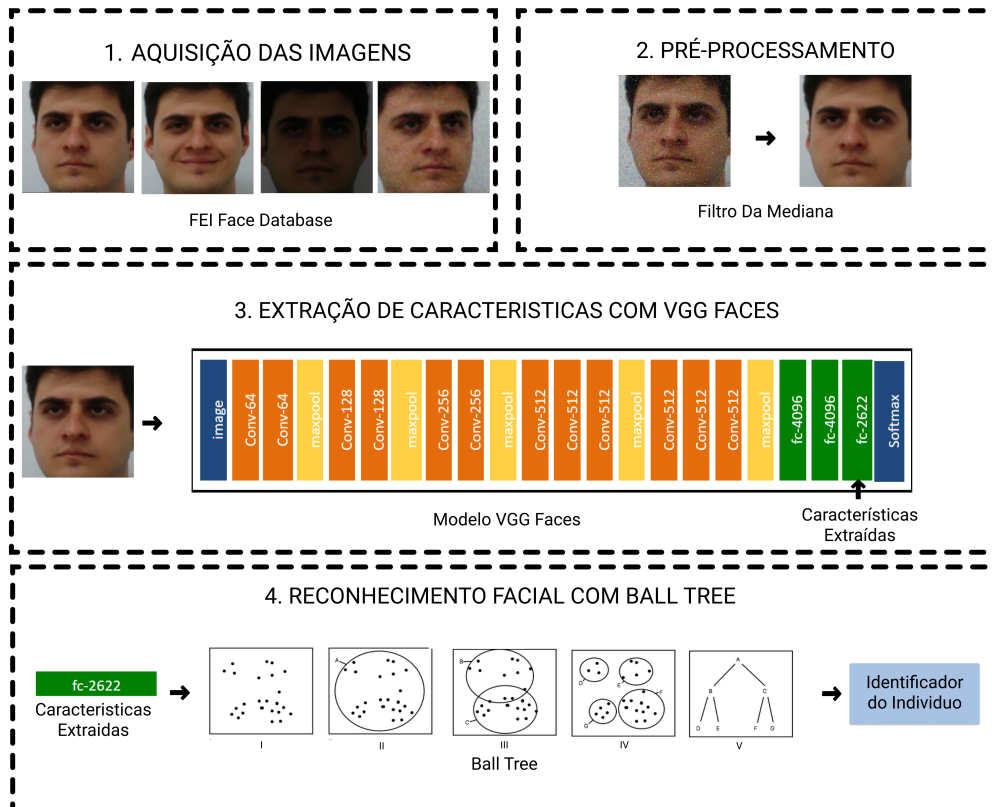


Figura 1. Fluxograma do método proposto dividido em quatro etapas.

3.1. Aquisição das imagens

Para avaliar o desempenho dos métodos de pré-processamento, foram selecionados quatro conjuntos de dados, são eles:

Base AT&T database of faces [AT&T 2002], composta de dez imagens diferentes de quarenta indivíduos distintos. Cada indivíduo possui dez fotos que foram tiradas em um ambiente com um fundo escuro homogêneo e em momentos diferentes, variando a iluminação, expressões faciais como olhos abertos ou fechados, sorrindo ou não e detalhes faciais como óculos ou sem óculos.

A segunda base é a FEI Face Database brasileiro [de Oliveira Junior and Thomaz 2006]. Ela é composta por 14 imagens de cada um dos 200 indivíduos distintos, somando um total de 2.800 imagens. Todas as imagens são

coloridas e tiradas contra um fundo branco homogêneo em uma posição frontal vertical com rotação de perfil, variando até no máxima de 180 graus.

O conjunto de dados CASIA-WebFace-cropped [NLPR-CBSR 2020], é um conjunto de imagens modificadas do banco de dados original CASSIA-WebFace. A sua modificação consiste em extrair e fornecer apenas as imagens das faces dos indivíduos presentes na base de dados original. O conjunto de dados contém 494.414 imagens de rosto de 10.575 identidades reais coletadas da web.

A base de dados UWA Hyperspectral Face Database [UWA 2013] é composta por imagens de rosto hiperespectral contendo cubos de imagem hiperespectral de 78 indivíduos obtidas em diferentes sessões. Essa base foi utilizada para simular testes com baixa iluminação. Exemplos de cada base de dados são mostrados na Figura 2.

E para ampliar a quantidade de tipos de ruídos testados, foram criadas novas bases de imagens adicionando ruídos artificiais às bases anteriormente mencionadas. Os ruídos usados são o ruído gaussiano e o ruído sal e pimenta. O ruído gaussiano consiste em um ruído estatístico que tem uma função de densidade e de probabilidade. Esse ruído se assemelha aos ruídos presentes em fotos com problemas de iluminação ou de alta temperatura durante a aquisição da imagem. O ruído sal consiste na distribuição de pixels corrompidos com valores ou muito altos ou muito baixos. Esses ruídos são naturalmente gerados por erros durante a transmissão da imagem, ocasionando o rompimento de pixels.

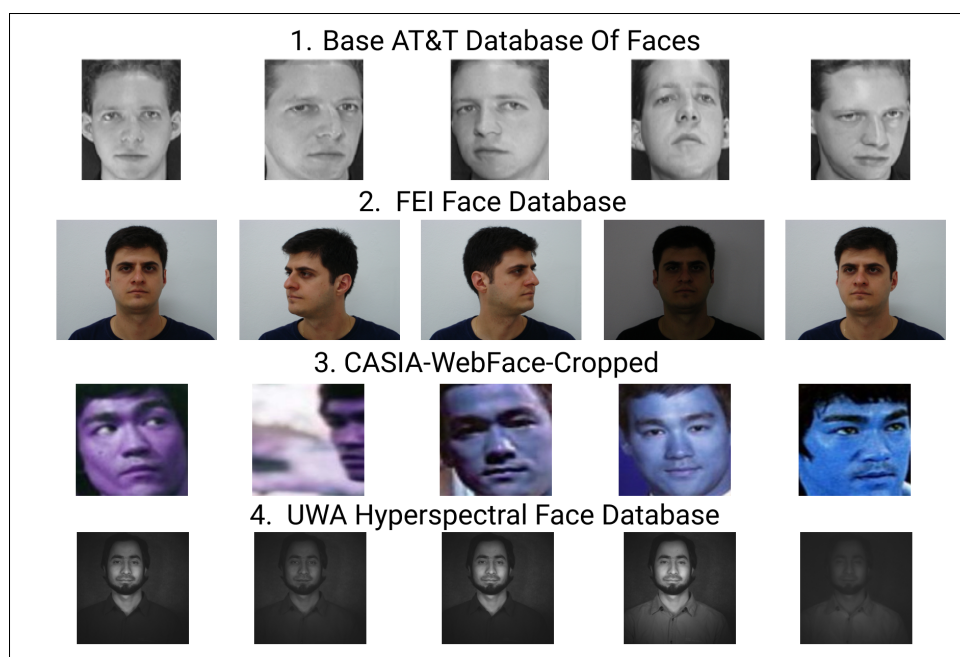


Figura 2. Exemplos de imagens presentes nas bases de dados.

3.2. Pré-Processamento

Neste trabalho foram testados os métodos de processamento escala de cinza, filtro gaussiano, filtro da mediana, filtro linear, equalização de histograma e transformação logarítmica. Para proporcionar novos testes que simulassem uma baixa qualidade referente ao dispositivo que realiza a captura das fotos, foram inseridos ruídos artificiais (ruído gaussiano e ruído sal e pimenta).

Escala de cinza: esse método consiste em converter as imagens para escala de cinza. Seu objetivo é de realçar características das fotos que não estão facilmente visíveis em escala RGB.

Filtro gaussiano: também conhecido como filtro de distribuição normal, possui a função de realce da imagem. A filtragem consiste no cálculo do valor do pixel central da máscara na imagem. A função utiliza os parâmetros média e desvio padrão obtidos usando o conjunto de pixels da área coberta pela máscara [Jesus and Jr 2015].

Filtro da mediana: é um método de transformação bastante usado para suavizar ruídos do tipo impulsivo. Sua filtragem consiste em atribuir o valor da mediana ao pixel central, usando os valores do conjunto de pixels da área coberta pela máscara [Jain 1995].

Filtro linear: Esse método possibilita a suavização que realça detalhes da imagem e minimizam efeitos de ruído, sem alterar a média da imagem [Jain 1995].

Equalização de histograma: Função de mapeamento baseado na função de distribuição acumulada. A imagem resultante possui valores do histograma próximos a uma distribuição uniforme, realçando-a, já que os seus pixels são distribuídos de forma a preencher o máximo do intervalo da faixa de níveis de cinza da imagem [Kim et al. 2001].

Transformações logarítmicas de intensidade: Método que realiza o mapeamento de uma determinada faixa de valores, aumentando o contraste das áreas com baixas intensidades e diminuindo o contraste das áreas com altas intensidades [Savvides and Kumar 2003].

3.3. Extração de características

Nesta etapa foi realizada a extração de características utilizando o modelo de rede neural convolucional (CNN) Vgg Faces. Essa rede é composta por 22 camadas e 37 unidades de profundidade. O descritor VGG-Face é calculado usando uma implementação de CNN com base na arquitetura VGG-Very-Deep-16. Para esta abordagem, foi utilizado pesos pré-treinados, pela base de dados VGG-Faces dataset, que é composta por imagens de 2.622 celebridades distintas. Para a extração de características das imagens, foi removida a última camada denominada de SoftMax, obtendo assim a representação da imagem, um vetor bidimensional de 2.622 características [Parkhi et al. 2015].

3.4. Reconhecimento

Para o reconhecimento foi utilizado o método Ball Tree, que consiste em uma árvore de pesquisa eficiente que é utilizada para consultas com a distância euclidiana. Cada nó desta árvore representa uma bola, ou seja, uma esfera que contém um subconjunto dos pontos que serão posteriormente pesquisados. Para classificar o indivíduo testado, foi selecionada a representação com a menor distância em relação a imagem de consulta.

4. Resultados e discussão

Para as bases de dados FEI, CASIA-WebFace-cropped e UWA Hyperspectral, por serem bancos de dados muito extensos, foi extraída uma amostragem de indivíduos que possibilitasse os testes. Sendo assim a quantidade total de imagens em cada uma das bases foi a seguinte: Base AT&T com 246 imagens, FEI Face Database com 1200 imagens, CASIA-WebFace-cropped com 1200 imagens e UWA Hyperspectral Face Database com

468 imagens. Em cada banco de imagens foi utilizada uma proporção de 50% para treino e 50% para teste. A métrica de avaliação utilizada foi a acurácia, que representa a divisão do número de acertos pela quantidade de predições realizadas.

Tabela 1. Resultados obtidos
Base de dados sem ruído artificiais

Pré-Processamento	CASIA-WebFace	AT&T	FEI	UWA Hyperspectral
Original	38	98	82	99
Escala de Cinza	35↓	99,1↑	84↑	96,1↓
Gauss.	37↓	97↓	82	100↑
Median Blur	37↓	95↓	81↓	98↓
Linear	36↓	96↓	82	99
Equali. Histogram	32↓	100↑	80↓	97↓
Transfo. Logarít.	28↓	100↑	71↓	100↑
Base de dados com ruído Gaussiano				
Pré-Processamento	CASIA-WebFace	AT&T	FEI	UWA Hyperspectral
Original	32	93	79	85
Escala de Cinza	27↓	96↑	76↓	77↓
Gaussiano	33↑	94↑	77↓	79↓
Filtro da mediana	32	93	78↓	63↓
Linear	34↑	92↓	77↓	87↑
Equali. Histogram	29↓	93	74↓	76↓
Transfo. Logarít.	18↓	98↑	71↓	80↓
Base de dados com ruídos Sal-e-Pimenta				
Pré-Processamento	CASIA-WebFace	AT&T	FEI	UWA Hyperspectral
Original	30	90	77	64
Escala de Cinza	30	93↑	75↓	51↓
Gaussiano	30	90	77↑	73↑
Filtro da mediana	37↑	94↑	81↑	99↑
Linear	31↑	96↑	76↓	82↑
Equali. Histogram	30	93↑	74↓	97↑
Transfo. Logarít.	9↓	46↓	70↓	53↓
Base de dados com ruídos Gaussiano e Sal-e-Pimenta				
Pré-Processamento	CASIA-WebFace	AT&T	FEI	UWA Hyperspectral
Original	28	83	75	37
Escala de Cinza	27↓	84↑	70↓	34↓
Gaussiano	27↓	93↑	77↑	75↑
Filtro da mediana	32↑	93↑	78↑	58↑
Linear	28	85↑	75	61↑
Equali. Histogram	25↓	83	70↓	66↑
Transfo. Logarít.	8↓	58↓	57↓	64↑

Na Tabela 1 são apresentados os resultados obtidos. A primeira parte da tabela mostra os resultados obtidos nas imagens originais, já na segunda parte são mostrados os resultados após a adição de ruído gaussiano. Na terceira e quarta são mostrados os resultados após a adição de ruídos Sal e Pimenta, e após a junção dos dois ruídos mencionados

anteriormente. Para melhorar a análise das tabelas, suas linhas foram destacadas, quando houver efeito negativo (representado pelo símbolo “↓”) e positivo (representado pelo símbolo “↑”). Também foi destacado em negrito os melhores resultados para cada base.

A base CASIA-WebFace foi a que obteve o menor desempenho dentre as avaliadas, nem mesmo a etapa de pré-processamento melhorou os resultados. O principal motivo para esse desempenho de acurácia abaixo de 40%, é relacionado a própria base, pois ela é bem desafiadora e além disso, algumas imagens foram rotuladas automaticamente e portanto estão rotuladas incorretamente. Ou seja, essa base de dados necessita de um pré-processamento amostral, não abordado neste trabalho.

Já na base AT&T, foram obtidos os melhores resultados, alcançando um percentual de acerto de 100% nos testes usando equalização de histograma e transformações logarítmicas. O alto desempenho pode ser justificado pela menor complexidade dessa base, pois as imagens foram obtidas em ambientes com iluminação controlada. Esses resultados mostram que realizando apenas o realce das características dos indivíduos é suficiente para uma melhoria nos resultados.

Na base FEI, foi obtido o melhor resultado quando as imagens foram convertidas para nível de cinza, obtendo um percentual de acerto de 84%. Ao analisar as imagens usadas nos testes, foi possível constatar que a amostra que representa o percentual de erros de predição, é composta por imagens com o rosto em perfil ou com mínima iluminação.

A base UWA Hyperspectral também obteve melhores resultados, com percentual de acerto de 100% usando filtro gaussiano e equalização de histograma. Por ser uma base com apenas características adversativas de iluminação, é possível concluir que os métodos de pré-processamento que produziram os melhores resultados são eficientes para o tratamento do problema de iluminação.

Pela análise das tabelas é possível concluir que os resultados com adição de ruído foram inferiores aos testes nas imagens sem ruído. Também foi possível concluir que é positivo para o desempenho do sistema de reconhecimento aplicar o filtro da mediana para todas as 4 bases, quando inserimos o ruído Sal e Pimenta, como mostrada na Figura 3.

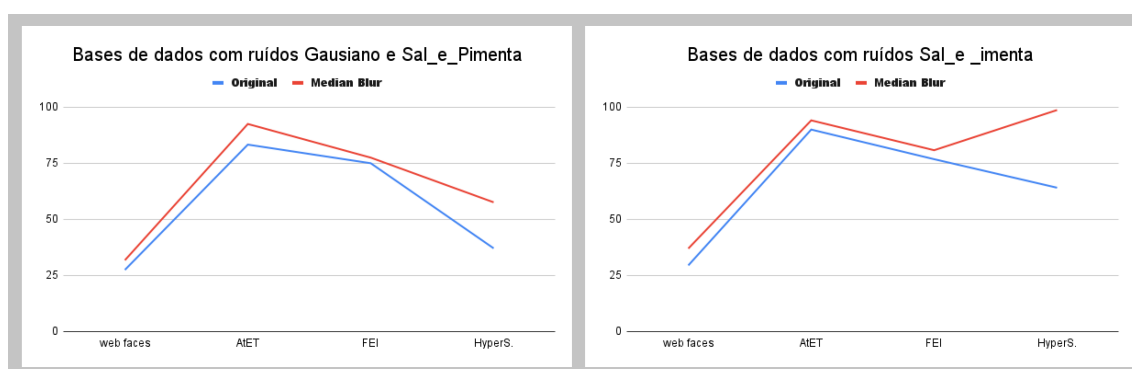


Figura 3. Resultados do filtro da mediana.

5. Conclusão

Este trabalho apresenta uma análise de técnicas de pré-processamento aplicados em sistema de reconhecimento facial. De acordo com os resultados obtidos, foi possível ob-

servar que em casos onde a base possui problema de iluminação, o realce de características por equalização do histograma e transformação logarítmica foram mais eficientes. Também foi possível concluir que o uso do filtro da mediana melhora o desempenho em situações onde as imagens possuem ruídos do tipo Sal-e-Pimenta. É importante destacar que a principal contribuição do trabalho foi a busca das melhores circunstâncias para a utilização dos métodos de pré-processamento. É esperado que em trabalhos futuros, sejam encontrados melhores descritores e classificadores, como também melhores parâmetros e métodos de pré-processamento de imagem.

Referências

- AT&T (2002). The at&t database of faces. <https://www.kaggle.com/kasikrit/att-database-of-faces>. Acesso em: 09/01/2021.
- Bah, S. M. and Ming, F. (2020). An improved face recognition algorithm and its application in attendance management system. *Array*, 5:100014.
- de Oliveira Junior, L. L. and Thomaz, C. E. (2006). Fei face database. <https://fei.edu.br/~cet/facedatabase.html>. Acesso em: 09/03/2021.
- Han, H., Shan, S., Chen, X., and Gao, W. (2013). A comparative study on illumination preprocessing in face recognition. *Pattern Recognition*, 46(6):1691–1699.
- Jain, R. (1995). *Machine Vision*. McGraw-Hill.
- Jesus, E. O. and Jr, R. C. (2015). A utilização de filtros gaussianos na análise de imagens digitais. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 3.
- Kim, J.-Y., Kim, L.-S., and Hwang, S.-H. (2001). An advanced contrast enhancement using partially overlapped sub-block histogram equalization. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(4):475–484.
- Lee, H., Park, S.-H., Yoo, J.-H., Jung, S.-H., and Huh, J.-H. (2020). Face recognition at a distance for a stand-alone access control system. *Sensors*, 20(3).
- Luo, M., Cao, J., Ma, X., Zhang, X., and He, R. (2021). Fa-gan: Face augmentation gan for deformation-invariant face recognition. *IEEE Transactions on Information Forensics and Security*, 16:2341–2355.
- NLPR-CBSR (2020). Casia-webface cropped. <https://www.kaggle.com/unkownhihi/casiawebface-cropped>. Acesso em: 09/03/2021.
- Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). *Deep face recognition*. British Machine Vision Association.
- Savvides, M. and Kumar, B. V. (2003). Illumination normalization using logarithm transforms for face authentication. In Kittler, J. and Nixon, M. S., editors, *Audio- and Video-Based Biometric Person Authentication*, pages 549–556, Berlin, Heidelberg. Springer Berlin Heidelberg.
- UWA (2013). Uwa hyperspectral face database. <https://openremotesensing.net/knowledgebase/uwa-hyperspectral-face-database-tip-2015-and-bmvc-2013/>. Acesso em: 09/03/2021.