

Classificação do óleo de algodão em processo de semirrefino usando técnicas de processamento de imagem

Patrícia Dias Araújo¹, Maria Victória Mera Moya² e Iális Cavalcante de Paula Junior³.

Resumo— O presente artigo apresenta um modelo de classificação nas etapas do processo de semirrefino do óleo de algodão por processamento de imagens. Foi obtido um sistema simples de análise eficiente, possibilitando sua utilização em chão fabril, auxiliando assim no processo quanto à assertividade no estado do semirrefino desejado e seu controle de qualidade. Foram utilizados dois algoritmos diferentes de reconhecimento de padrões: *K-Nearest Neighbors* e *Quadratic Discriminant Analysis*, tendo como validação a estratégia de *Leave-one-out*. Os dois algoritmos conseguiram taxas de acerto de mais de 90%, sendo o QDA o de melhor resultado com mais de 94% de acerto.

Palavras-Chave— *Classificação de óleo, óleo de Algodão, Processamento de Imagens, KNN, QDA, Matriz de Confusão, Leave-one-out.*

Abstract— The present article presents a classification model of steps in the process of semi - refining of cotton oil by image processing. A simple system of efficient analysis was obtained, allowing its use on the factory floor, thus assisting in the process as to the assertiveness in the desired semi-refined state and its quality control. Two different pattern recognition algorithms were used: *K-Nearest Neighbor* and *Quadratic Discriminant Analysis*, with validation of the *Leave-one-out* strategy. Both algorithms achieved hit rates of more than 90%, with QDA being the best result with more than 94% accuracy.

Keywords— *Oil Classification, Cotton Oil, Image Processing, KNN, QDA, Matrix of Confusion, Leave-one-out.*

I. INTRODUÇÃO

A procura do óleo vegetal tanto para fins alimentícios como para uso de matéria prima na produção de biodiesel tem se tornado cada vez maior. Um exemplo é o óleo de algodão, qual é frequentemente utilizado como medida padrão para a qualidade de outros óleos. Este apresenta características que atendem aos requisitos de quase qualquer aplicação de alimentos imagináveis. O óleo de algodão possui um sabor e odor suave, já seu grau de tonalidade de cor depende do nível do processo de refino como a maioria dos óleos [1].

O processo de refino pode ser definido como um conjunto de métodos que visam transformar o óleo bruto em óleo comestível, embora existam casos de consumo de óleos brutos, como o azeite de oliva e o de dendê. A finalidade da refinação é obter melhorias na aparência, no odor e sabor do óleo, por meio da remoção dos seguintes componentes:

- a) Substâncias coloidais, proteínas, fosfatídeos e produtos de sua decomposição;
- b) Ácidos graxos livres e seus sais, ácidos graxos oxidados, lactonas, acetais e polímeros;
- c) Substâncias coloridas como clorofila, xantofila, carotenoides, incluindo-se neste caso o caroteno ou provitamina A;
- d) Substâncias voláteis como hidrocarbonetos, álcoois, aldeídos, cetonas e ésteres de baixo peso molecular;
- e) Substâncias inorgânicas como os sais de cálcio e de outros metais, silicatos, fosfatos, dentre outros minerais;
- f) Umidade [2].

Atualmente o óleo de algodão (*Gossypium hirsutum latifolium* Hutch LR) é considerado uma das principais fontes de biomassa para produção do biodiesel no Brasil. Com isso tem se o aumento da relevância no controle de qualidade da produção tanto em relação à matéria prima como ao produto final. Veras et al. (2012) desenvolveram metodologias analíticas utilizando as espectroscopias no infravermelho próximo NIR (*Near Infrared Spectroscopy*) e no ultravioleta-visível (UV-vis) para classificar amostras de biodiesel a partir do seu óleo de origem, utilizando como técnica de reconhecimento de padrões a Modelagem Independente e Flexível por Analogia de Classe SIMCA (*Soft Independent Modeling by Class Analogy*) [3].

Balabin e Safieva (2011) também desenvolveram técnicas de classificação de óleo utilizando espectroscopia NIR e técnicas de classificação multivariadas como Análise Discriminante pelos Mínimos Quadrados Parciais PLS-DA (*Partial Least Square Discriminant Analysis*) [4]. No entanto, estas técnicas possuem custos elevados de instrumentos e de manutenção, o que se faz necessário o desenvolvimento de uma metodologia alternativa que seja simples, de baixo custo, fácil implementação no processo fabril e que classifique em qual etapa do processo de refino ou semirrefino o óleo se encontra. Tendo como fator positivo o uso da imagem digital, substituindo o sistema visual humano e eliminando o caráter subjetivo da análise, o que poderia induzir a erros.

Portanto, o presente trabalho a partir do processamento de imagens utilizando algoritmos de reconhecimento de padrões, tem como objetivo apresentar uma técnica simples e de baixo custo que realize a classificação dos diferentes estados do processo de semirrefino do óleo de algodão. Para isso foram

Patrícia Dias Araújo, Maria Victória Mera Moya e Iális Cavalcante de Paula Junior, Universidade Federal do Ceará (UFC), Sobral - CE, Brasil, E-mails: patriciadiasaraujo.18@gmail.com, vicky_memo@hotmail.com e ialis@sobral.ufc.br
Este trabalho foi parcialmente financiado pela CAPES.

utilizados dois tipos de classificadores para comparação de dados, o *K-Nearest Neighbor (KNN)* e *Quadratic Discriminant Analysis (QDA)*, sendo aplicada como estratégia de validação a *Leave-one-out*. A base de dados utilizada foi de 24 imagens, divididas em 4 subimagens cada uma, obtendo no total 96 subimagens. O ambiente empregado para captura das imagens foi o laboratório de uma fábrica de derivados do caroço de algodão, localizada no interior do Ceará.

O presente artigo está organizado nas seguintes seções: a seção II aborda a teoria relacionada ao trabalho, explicando o processo e a importância da diferenciação do óleo a partir da cor no processo de semirrefino; na seção III é apresentado o método proposto, que inclui técnicas de processamento de imagens e reconhecimento de padrões, a seção IV contém os resultados obtidos no processo e discussões, e finalmente na seção V a conclusão do trabalho.

II. PROCESSO DE SEMIRREFINO DO ÓLEO DE ALGODÃO

O óleo da semente de algodão é o mais antigo óleo vegetal produzido industrialmente e consumido em larga escala no Brasil [5]. Esse óleo, em estado bruto, apresenta uma coloração bastante intensa, que quando utilizado para fins alimentícios passa por um processo de clareamento que pode ser o processo de refino ou o semirrefino, onde a diferença está na etapa de desodorização. Durante o refino, a clarificação é considerada como uma etapa de grande importância na determinação da qualidade e estabilidade do produto final.

O processo de semirrefino do óleo de algodão analisado neste trabalho possui seis etapas, as quais são apresentadas na Tabela I.

TABELA I. PROCESSO DE SEMIRREFINO DO ÓLEO DE ALGODÃO

Estado do óleo durante o processo (cor)	Características
	B -Inicialmente tem-se o óleo bruto o qual após ser extraído por processo de prensagem é destinado para centrifugas para o início do processo de semirrefino.
	1R -Processo de primeiro refino: é dosada a soda caustica, a temperatura é elevada para em torno de 65° C e o estado é mantido por um tempo de retenção.
	2R -Segundo refino, onde é realizada uma pré-lavagem, elevada a temperatura até 70°C e não há tempo de retenção.
	L -Lavagem do óleo, onde é acrescentado cerca de 20% de água em relação ao volume de óleo e a temperatura é elevada a 80°C. Após é retirada o máximo possível de água pelo o processo de centrifugação.
	LS -Secagem do óleo, qual é realizada por um secador a vácuo, retirando assim o excesso de água que permaneceu do processo anterior.
	BR -Branqueamento, onde ocorre a adição de um auxiliar filtrante ao óleo, sendo retido após em um clarificador á 110°C e só então filtrado.

Quando o refino é realizado eficientemente, o óleo obtido é quase incolor, atendendo a uma das principais exigências do consumidor. Dessa forma, além de possibilitar inferências sobre a qualidade do processamento do óleo vegetal, a cor também pode ser utilizada para avaliar seu estado de conservação. Uma vez que iniciado o processo de deterioração e alterações organolépticas ocorre o aparecimento de cor e outras alterações no óleo [6,7].

A. A importância da cor na classificação do óleo

Para análise laboratorial do óleo vegetal são analisadas as cores primárias R (vermelho), Y (amarelo), B (azul) e N (neutro). O conjunto destas variáveis indica a predominância de cor do óleo e conseqüentemente podem indicar os componentes que se encontram em maior concentração, ou seja, valores altos de leitura R (vermelho) e ou Y (amarelo) indicam altas concentrações de carotenóides. No caso de óleos refinados existe um padrão na legislação que define que todos os óleos, independente da origem, devem estar dentro deste limite, uma vez que aparência do óleo é uma das primeiras barreiras para a aceitação do produto pelos consumidores [8]. A característica de cor é uma das propriedades medidas para avaliar se a etapa de clarificação durante o semirrefino foi adequada e se o óleo produzido atende aos requisitos da legislação. Quanto mais intensa for a cor do óleo bruto, maior deverá ser o controle do processo [9].

III. MÉTODO PROPOSTO

O presente artigo possui a estrutura apresentada no fluxograma abaixo, onde cada bloco tem como finalidade processar a imagem de entrada e repassar seu resultado para o próximo bloco. Nas seções a seguir, a função de cada bloco será explicada com detalhes. Para a execução do método, foram utilizadas as ferramentas do *software* Matlab. A Figura 1 apresenta a divisão do algoritmo de classificação do processo de semirrefino do óleo em diagrama de blocos.

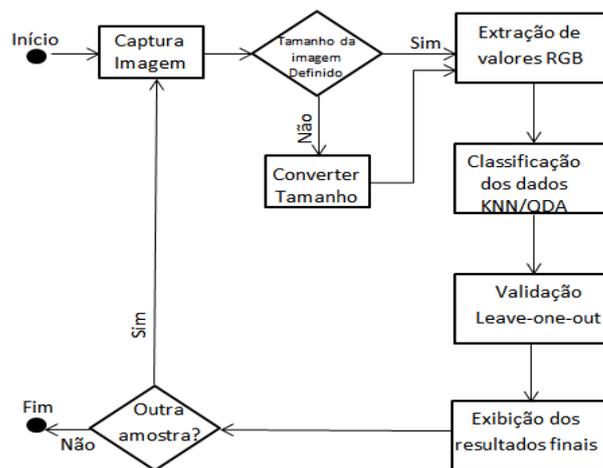


Fig. 1. Diagrama de blocos algoritmo de classificação do processo de semirrefino do óleo de algodão. Fonte: Produzido pelo Autor

A. Aquisição de Imagens e Pré-Processamento

Com o intuito de que as imagens a serem analisadas fossem obtidas nas mesmas circunstâncias, um sistema para captura foi montado. Neste sistema foi usado uma webcam *Microsoft LifeCam HD-3000* conectada a um *notebook* e fixada a uma altura aproximada de 28,5 centímetros das amostras de óleo. Foi definido também um padrão de volume,

posição e características do recipiente com óleo; o material do recipiente utilizado era aço inoxidável para amenizar interferências na cor do óleo. A Figura 2 apresenta o sistema utilizado na aquisição das imagens.



Fig. 2. Sistema para captura das imagens. Fonte: Produzida pelo autor

Após a aquisição das imagens de dimensão 1280x800 pixels, foi utilizado um algoritmo que identifica a área válida para análises, ou seja, salva numa nova imagem o espaço que não tem contato com elementos alheios ao óleo, por exemplo, as bordas do recipiente ou a mesa. Desta área válida obtida, foram geradas quatro subimagens de dimensões 100x100 pixels em posições específicas. A Figura 3 apresenta um exemplo da imagem original obtida pelo sistema de aquisição e suas subimagens.

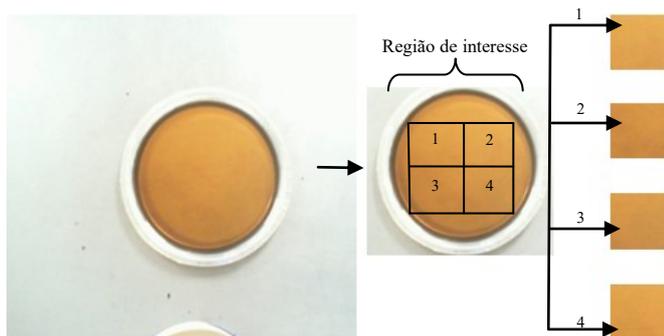


Fig. 3. Subimagens do processo de Segundo refino. Fonte: Produzida pelo autor

O segundo passo foi adotar os modelos de cor RGB, CMY, HSV, YCBCR e YIQ para obtenção de atributos que serão utilizados como informação analítica. O modelo RGB (*red, green, blue*) é um modelo linear que não consegue representar perfeitamente o modo com que o olho humano percebe as cores; é um modelo orientado para hardware, utilizado para monitores coloridos e uma ampla classe de câmeras de vídeo a cores; o modelo CMY (*cyan, magenta, yellow*) e o modelo CMYK (*cyan, magenta, yellow, black*) são modelos para impressão colorida; o modelo HSI (*hue, saturation, intensity*) que corresponde estreitamente à forma como os seres humanos descrevem e interpretam as cores, separando o componente intensidade das informações de cores (matriz e saturação) em uma imagem colorida [10]; e o modelo YIQ se baseia na separação dos sinais de cor RGB em um sinal de luminosidade ou luminância (Y) e dois sinais de cromaticidade ou diferença de cor (IQ) [11].

A quantidade de atributos utilizados para classificação do óleo depende do modelo de cor adotado e a média dos valores de cada faixa de cor, por exemplo, no caso do modelo RGB,

foram contidos três atributos com a informação da média dos valores de cada faixa.

B. Classificação de amostras por KNN e QDA

O Algoritmo KNN é um classificador de aprendizagem indutiva “*lazy-learning*”, não linear e nem paramétrico, muito utilizado para estimação estatística e reconhecimento de padrões. Conta com uma base de dados utilizada para assimilar a classe correspondente a uma amostra desconhecida através de comparações [12]. A comparação é realizada pelo método da distância mínima, atribuindo a cada elemento desconhecido à classe dos K vizinhos mais próximos [13]. O método KNN pode utilizar diferentes distâncias para fazer o cálculo da classificação, entre elas estão as distâncias de Manhattan, Mahalanobis e a Euclidiana [14].

A regra de classificação do KNN simplesmente retém todo o conjunto de treinamento durante a aprendizagem e atribui a cada amostra analisada a sua classe pertencente, sendo essa representada pelo rótulo majoritário de seus K vizinhos mais próximos no conjunto de treinamento [12]. No presente artigo o valor de K, que determina a quantidade de vizinhos que serão avaliados para classificar o indivíduo, foi obtido pelo o método de tentativas e erro. Logo, K = 4 uma vez que com esse valor se obteve os melhores resultados.

Segundo Regazzi (2001), embora a distância euclidiana seja uma medida de dissimilaridade, essa pode ser utilizada como uma medida de semelhança, pois quanto maior seu valor, menos parecidos são os indivíduos ou unidades mostrais [15]. No presente estudo foi adotada a distância Euclidiana, onde ao fornecer a distância direta entre os pontos em análise, foi obtida a semelhança entre os mesmos.

A técnica QDA é utilizada para problemas de classificação de observações quando as regiões de classificação são definidas por funções quadráticas. Tem por objetivo encontrar uma combinação quadrática de todas as variáveis de forma que o valor resultante dessa combinação, chamado discriminante, tenha valores muito diferentes para observações pertencentes a classes diferentes e valores próximos para observações da mesma classe [16]. A técnica QDA se baseia em duas hipóteses:

1. Todas as populações, que dão origem aos grupos de dados, possuem curva de distribuição normal;
2. As matrizes de covariância de cada grupo são diferentes.

A técnica se baseia simplesmente na maximização das probabilidades posteriores, $\Pr(K=k | X=x)$ ou seja, a probabilidade de uma observação pertencer a um grupo k, dado que ela é definida por x [17].

C. Validação por Leave-one-out

O *Leave-one-out* é um método de validação cruzada, apresentando o caso extremo da estratégia *K-fold*, em que o valor de K é a quantidade total de amostras. Desse modo, o algoritmo executa N iterações, e em cada uma delas o conjunto de teste está formado por uma única amostra e todas as outras amostras formam o conjunto de treinamento. Ao subtrair o valor verdadeiro com o valor predito tem-se o erro da estimação de cada ponto para cada método usado [18].

IV. RESULTADOS

Na Tabela II são apresentados os percentuais de acerto na classificação das seis etapas do processo de semirrefino do óleo de algodão. Com os classificadores KNN e QDA para os diferentes modelos de cores adotados.

TABELA II. ACURÁCIA MÉDIA CONFORME MODELO DE COR E CLASSIFICADOR

Acurácia Média (%)	Modelo de Cor	KNN	QDA
	RGB	90,6	94,7
	CMY	90,0	94,7
	HSV	86,4	93,7
	YCBCR	92,7	93,7
	YIQ	92,7	94,7

Os modelos que apresentaram os melhores resultados na classificação tanto para o algoritmo de KNN e QDA foram os modelos de cor RGB e YIQ. As matrizes de confusão para a classificação KNN nos modelos de cor RGB e YIQ são apresentadas na Tabela III e IV respectivamente. As entradas diagonais neste tipo de matriz mostram o número de acertos na classificação. Na Tabela III as classes de óleo bruto (B) e primeiro refino (1R) tiveram 100% de acerto, já as classes de óleo lavado (L) e lavado e seco (LS) foram os que apresentaram menor percentual de acerto, tendo em vista que essas duas classes possuem características semelhantes. Já no modelo de cor YIQ além das classes de óleo bruto (B) e primeiro refino (1R) a classe de óleo lavado (L) também apresentou 100% de acerto.

TABELA III. MATRIZ DE CONFUSÃO PARA KNN; K= 4; MÉTODO DE VALIDAÇÃO LEAVE-ONE-OUT; MODELO DE COR RGB

Classe do Óleo	Classes Atribuídas						Acerto (%)
	B	BR	LS	L	1R	2R	
B	16	0	0	0	0	0	100
BR	0	15	1	0	0	0	93.75
LS	0	1	11	4	0	0	68.75
L	0	0	1	14	0	1	87.5
1R	0	0	0	0	16	0	100
2R	0	0	0	1	0	15	93.75

TABELA IV. MATRIZ DE CONFUSÃO PARA KNN; K= 4; MÉTODO DE VALIDAÇÃO LEAVE-ONE-OUT; MODELO DE COR YIQ

Classe do Óleo	Classes Atribuídas						Acerto (%)
	B	BR	LS	L	1R	2R	
B	16	0	0	0	0	0	100
BR	0	15	1	0	0	0	93.75
LS	0	1	11	4	0	0	68.75
L	0	0	0	16	0	0	100
1R	0	0	0	0	16	0	100
2R	0	0	0	1	0	15	93.75

Para a classificação com QDA a matriz de confusão apresentou os mesmos resultados tanto para o modelo de cor RGB como para o modelo YIQ, os quais são apresentados na Tabela V.

TABELA V. MATRIZ DE CONFUSÃO PARA O CLASSIFICADOR QDA COM MÉTODO DE VALIDAÇÃO LEAVE-ONE-OUT

Classe do Óleo	Classes Atribuídas						Acerto (%)
	B	BR	LS	L	1R	2R	
B	16	0	0	0	0	0	100
BR	0	15	1	0	0	0	93.75
LS	0	0	13	2	0	1	81.25
L	0	0	0	16	0	0	100
1R	0	0	0	0	16	0	100
2R	0	0	1	0	0	15	93.75

Os melhores resultados obtidos para a classificação do óleo vegetal foram com o uso do classificador QDA com o modelo de cor RGB. Segundo os dados experimentais do óleo, a cor de maior importância é a vermelha devido a presença de carotenóides. Os resultados da classificação confirmam esta predominância, ao obter como melhor modelo de classificação o RGB com análises da faixa vermelha.

V. CONCLUSÕES

Este artigo busca desenvolver um sistema que substitua a atual técnica manual de classificação nas etapas do processo de semirrefino do óleo de algodão. Os resultados obtidos apresentaram um método semiautomático, mais eficiente e confiável, mostrando que as técnicas de processamento e reconhecimento de padrões apresentadas são capazes de extrair de uma imagem da amostra de óleo, informações suficientes para definir em que etapa do processo ele se encontra. O classificador QDA apresentou as melhores respostas com 94,7% de acerto na classificação do óleo, o que indica um bom método para realização da classificação do processo de semirrefino. Logo os resultados obtidos na classificação foram satisfatórios cumprindo-se assim o objetivo inicial do artigo.

AGRADECIMENTOS

A Indústria e Comércio de Rações Golfinho, que disponibilizou do seu espaço e informações do seu processo para a realização da pesquisa aqui apresentada, ao PPGEEC-UFC e a CAPES.

REFERÊNCIAS

- [1] Dinesh K. Agarwal, Phundan Singh, Mukta Chakrabarty, A J Shaikh, S G Gayal, "Cottonseed Oil Quality, Utilization And Processing", Central Institute for Cotton Research, Nagpur, CICR TECHNICAL BULLETIN NO: 25, 2003.
- [2] Gontijo, J. M.; Hirakuri, M. M. H.; A. C. R. *Tecnologia para produção do óleo de soja*: descrição das etapas, equipamentos, produtos e subprodutos. 2ª Edição. ed. Londrina: [s.n.], vol. 2176-2937, 2015.
- [3] Costa, G. B. D. "Classificação Multivariada De Biodiesel B100 E B5 Usando Imagens Digitais". Universidade Estadual Da Paraíba. Campina Granda - PB. 2015.
- [4] Balabin, R.M.; Safieva, R.Z. Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data, *Analytica Chimica Acta*, vol.89, n.1, p. 190-197, 2011.
- [5] Cardello, H. M. A. B. et al. "Classificação Convencional de Óleo de Semente de Agodão (*Gossipium Hirsutum*)". UNICAMP. São Paulo, p. 77-87. 1995.
- [6] Oretto, E.; Fett, R. *Tecnologia de Óleos e Gorduras Vegetais na Indústria de Alimentos*. São Paulo: Varela Editora e Livraria, 1998.

- [7] Shahidi, F.; Zhong, Y. *Lipid Oxidation: Measurement Methods. In: Bailey's Industrial Oil and Fat Products*. 6th Edition. Six Volume. Canada: John Wiley & Sons, Inc., 8, 357-385, 2005.
- [8] Lawson, H. "Food oils and fats: technology, utilization and nutrition". New York: Chapman & Hall, 1995.
- [9] Silva, F. A. et al. "Aplicação do pré-processamento de imagens para otimização do reconhecimento de padrões na detecção de deficiência nutricional em espécies vegetais". Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR. João Pessoa. 2015.
- [10] Gonzalez, R.C.; Woods, R.E. "Processamento de imagens digitais". Tradução Roberto Marcondes Cesar Júnior, Luciano da Fontoura Costa. 1ª Ed. São Paulo: Editora Blucher, 2010.
- [11] Souza, L. Q. V. D.; Carneiro, L. I. S.; Candeias, A. L. B. "Modelos de Cores CMY e YIQ no SPRING e no MATLAB para Análise Visual de Imagens IKONOS e CBERS". *XIV Simpósio Brasileiro de Sensoriamento Remoto*, Natal, 2009. 7113-7180.
- [12] Bolandraftar, S. B. I. A. M. Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background. *S B Imandoust et al. Int. Journal of Engineering Research and Applications*, Tehran, 05 Sep-Oct 2013. 605 a 610.
- [13] Hodgson, M. E. "Reducing the computational requirements of the minimum-distance classifier". *Remote sensing of environment*, vol. 25, p. 117-128, 1988.
- [14] Silva, I. M. O. E. "Uso de Técnicas Computacionais Para O Reconhecimento de Áreas Sem Arborização a Partir de Imagens de Satélite". Centro Federal De Educação Tecnológica De Minas Gerais. Timóteo. 2014.
- [15] Regazzi, A. J. *INF 766 - Análise multivariada*. Viçosa: Universidade Federal de Viçosa, Centro de Ciências Exatas e Tecnológicas. Departamento de Informática, 2001. 166p. Apostila de disciplina.
- [16] Hastie et al., 2008 Hastie, T., Tibshirani, R., and Friedman, J. (2008). *The Elements of Statistical Learning*. Springer.
- [17] Paula, R. R. R. D. "Análise comparativa de modelos de estatística multivariada aplicados à previsão de níveis de poluentes atmosféricos". Escola Politécnica da Universidade de São Paulo. São Paulo. 2017.
- [18] Backes, A. R.; Sá Junior, J. J. D. M. "Introdução à Visão Computacional Usando Matlab". Rio de Janeiro: Alta Books, 2016.