

Relatório Final de Atividades

Análise e quantificação de danos em sementes de soja vinculado ao projeto Métodos para análise, caracterização e classificação de bioimagens.

Thullyo Radeli Castilho

Voluntário

Engenharia de Computação

Data de ingresso no programa: 10/2015

Prof(º). Dr(º). Fabrício Martins Lopes

Área do Conhecimento: 1.03.00.00-7 - ciência da computação

CAMPUS CORNÉLIO PROCÓPIO, 2016

THULLYO RADELI CASTILHO
FABRÍCIO MARTINS LOPES

RELATÓRIO FINAL DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA

Relatório Técnico do Programa de
Iniciação Científica e Tecnológica da
Universidade Tecnológica Federal do
Paraná.

RESUMO

No Brasil, a agricultura é uma das áreas que mais contribuem para a economia, sendo bastante reconhecida e figurando entre os maiores produtores em diversos setores, como no caso da soja, em que o país é o segundo maior produtor mundial. Devido a importância das culturas plantadas, sua produtividade, tanto em lavoura como para usos diversos, é alvo de regulações governamentais e de estudos contínuos, sempre procurando obter sementes com as melhores capacidades germinativas e produtivas o possível. Em especial, esta oleaginosa passa por diversos processos de beneficiamento antes de chegar à seus destinos de utilização final, tendo como um dos maiores objetivos, o de garantir uma qualidade mínima das sementes de soja, que está diretamente relacionada com sua produtividade. Devido à isso, além de regras estipuladas pelo governo, as empresas são livres para desenvolverem processos fabris adicionais próprios. Especificamente, a empresa Belagrícola S.A. possui um processo criado internamente para o controle e ajuste dos seus equipamentos de beneficiamento. Através de um estudo de caso nessa empresa a cerca deste controle próprio, foi possível constatar diversas deficiências em sua operação, principalmente devido à falta de automatização em determinadas etapas. Dentro deste contexto, o objetivo do presente trabalho foi desenvolver um método eficiente e automático para a classificação de defeitos em sementes de soja, através da aplicação dos conceitos de processamento digital de imagens e reconhecimento de padrões. Os resultados alcançados, onde se produziu um modelo de classificação capaz de obter uma acurácia média máxima de 81,25%, considerando as quatro classes de sementes estudadas, demonstram o potencial e a importância da pesquisa realizada.

Palavras-chave: Processamento de imagens, classificação, reconhecimento de padrões, defeitos em sementes de soja.

SUMÁRIO

INTRODUÇÃO	4
REVISÃO BIBLIOGRÁFICA	5
MATERIAIS E MÉTODOS	11
RESULTADOS E DISCUSSÕES	14
CONCLUSÕES	14
REFERÊNCIAS	15

INTRODUÇÃO

Devido aos avanços tecnológicos, atualmente, se tornou possível a produção e uso em grande escala das mais variadas formas de multimídia, como áudio, vídeos e imagens digitais [1]. Em especial, as imagens digitais constituem um importante meio pelo qual diversas aplicações são possíveis e realizadas, constituindo uma importante área de pesquisa chamada de Processamento de Imagens Digitais [2].

Dentre as aplicações envolvendo o processamento de imagens, uma de extrema importância é o reconhecimento de padrões de forma automática por algoritmos computacionais, principalmente devido às possibilidades e benefícios acarretados por essa abordagem, pois em muitos casos, quando tal trabalho é realizado por pessoas, o mesmo se torna ineficiente e custoso [3]. O espectro de aplicabilidade compreende desde o reconhecimento de simples caracteres até mesmo aplicações em agricultura [4, 5].

No Brasil, o agronegócio é um dos setores de maior importância para a economia, representando 23% do PIB em 2015 [6]. Na agricultura, dentre os cultivares produzidos, a soja é uma das mais proeminentes, atingindo aproximadamente 17 milhões de toneladas no Paraná, consistindo do segundo maior produtor brasileiro, e 96 milhões no país, o que faz do Brasil o segundo maior produtor mundial, atrás apenas dos Estados Unidos [7, 8].

As sementes de soja utilizadas para o plantio em larga escala, antes de chegar até o produtor rural, para que se garanta a maior taxa de germinação possível, devem passar por um processo de beneficiamento, passando por testes padronizados e regulados pelo governo [9]. O beneficiamento é composto de diversos processos que visam garantir a qualidade e capacidade de germinação dessas sementes e consiste em separá-las entre as de boa e má qualidade, de forma que apenas aquelas com melhor taxa de germinação sejam distribuídas para o plantio, sendo o restante redirecionado à outras aplicações, como produtos alimentícios e ração animal [10].

A taxa de germinação pode ser aferida de diversas formas, entre elas o teste tetrazólio, pêndulo ou até mesmo considerando características puramente visuais da semente, abrindo espaço para a aplicação de processamento de imagens também à esta área em específico [11].

Em especial, esse processo de beneficiamento ocorre normalmente por meios mecânicos, que consiste de processos automatizados, rotineiramente necessitando de calibração para que permaneçam eficientes e o mais precisos possível. Para a definição dos parâmetros das máquinas, utiliza-se de estatísticas de qualidade inferidas com base em amostras de grãos enquanto estes passam pelos processos de separação e beneficiamento, em que o registro de informações errôneas pode influenciar negativamente na qualidade final das sementes [12].

Assim, a base de motivação para este trabalho são as deficiências encontradas no processo de controle de qualidade, que foi desenvolvido pela própria empresa, para calibração das máquinas de beneficiamento e classificação de sementes de soja da empresa Belagrícola S.A., unidade de processamento de Tamarana (PR), as quais foram constatadas através de observações e encontros com a equipe responsável pelo processo fabril.

Neste contexto, pretende-se implementar um sistema baseado em processamento de digital de imagens, capaz de eliminar ou ao menos amenizar tais deficiências, proporcionando mais qualidade e confiabilidade ao processo de controle de qualidade através de sua automatização.

REVISÃO BIBLIOGRÁFICA

Defeitos em Sementes de Soja. Os defeitos, provocados por manipulação ou pragas, refletem fortemente na qualidade da semente, ocasionando, por exemplo, uma menor quantidade de nutrientes ou óleo quando processadas para produção de produtos diversos [12]. Reflete também no vigor, que por sua vez está ligado diretamente com a futura produtividade em lavoura, resultando no fato de que sementes defeituosas são de mais baixo valor [12, 13]. Há ainda de se considerar as implicações governamentais, pois é necessário garantir uma qualidade mínima das sementes distribuídas aos produtores rurais [9].

Assim, o pleno entendimento, a aplicação e o desenvolvimento de técnicas e equipamentos especiais para a detecção e processamento destes tipos de sementes defeituosas se faz de extrema importância, para que se cumpra a legislação vigente [12, 13].

Dano por Percevejo. Os percevejos pentatomídeos, são uma praga de extremo malefício ao cultivo da soja, eles se alimentam diretamente das sementes, cujos danos se refletem principalmente na qualidade destas [14, 10].

Quando estes insetos se alimentam da semente de soja, podem provocar um achatamento característico na região, além de comumente as inocularem com o fungo *Nematospora coryli* Peglion, o que causa necrose do local, onde se pode apresentar marcas típicas com um formato circular, de diferentes cores características, em que alguns casos ainda ocorrem lesões profundas, as quais podem não apresentar sinais visíveis no exterior [14, 10]. Alguns exemplos de sementes com a presença de danos por percevejo podem ser observados na Figura 1.



Figura 1. Exemplos de sementes com dano por percevejo. Os danos se caracterizam principalmente pela presença de pequenas manchas e partes da semente com aspecto amassado. Fonte: Autoria própria.

Semente com Esverdeado. A cor verde característica, presente em sementes imaturas, é devido a não deterioração da clorofila durante seu processo de maturação, o qual é resultado de plantas que passam por estresses bióticos ou abióticos, precipitando sua morte [15].

As principais consequências para uma semente apresentar esverdeado, são a redução da qualidade de óleo e isolados proteicos, além de acentuada redução do seu vigor, o que resulta em baixa produtividade após seu plantio [15]. A Figura 2 apresenta amostras de sementes esverdeadas.



Figura 2. Amostras de sementes esverdeadas. Fonte: Autoria própria.

Dano Mecânico. Uma das causas pela perda da qualidade de germinação em sementes de soja são os danos mecânicos, os quais são geralmente provocados durante a colheita [12, 10].

Algumas das partes mais importantes das sementes são também as mais sensíveis, como o eixo embrionário, que é recoberto por uma película frágil, podendo assim, caso sofra algum tipo de dano mecânico, impactar na variabilidade genética, a qual está diretamente relacionada com a qualidade das sementes [12, 10]. Normalmente encontram-se desde danos visíveis, como rachaduras e partes quebradas, até profundos, em que muitas vezes a inspeção visual simples pode não ser suficiente para sua detecção. Algumas sementes afetadas por danos mecânicos são mostradas na Figura 3.



Figura 3. Exemplos de sementes com danos mecânicos. Fonte: Autoria própria.

Imagens digitais. Uma imagem digital pode ser representada como um vetor bidimensional $M \times N$, ou vários destes, sendo um para cada componente de cor, onde cada elemento do vetor é chamado de *pixel*, o qual é uma função $f(x, y)$ em que $0 \leq x \leq M-1$ e $0 \leq y \leq N-1$, sendo que o valor da função é o brilho ou intensidade deste *pixel*, que fica localizado no ponto de coordenadas x e y da matriz [2, 16]. Um exemplo desta representação pode ser observado na Figura 4.

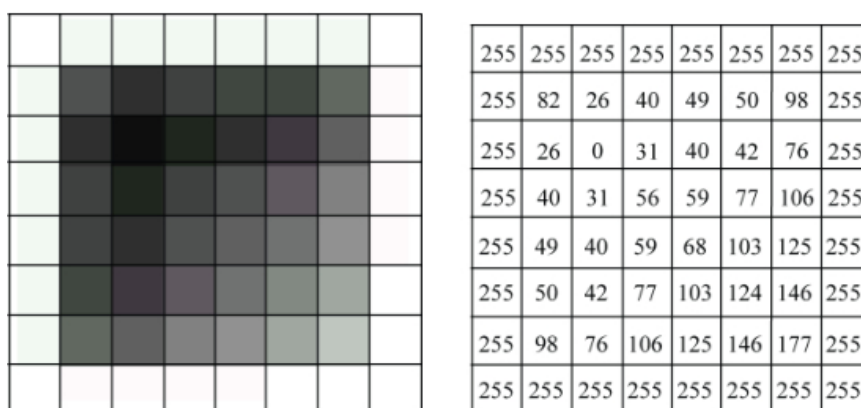


Figura 4. Exemplo de imagem digital em escala de cinza representada por uma matriz. Fonte: Adaptado de [17].

Caso a imagem seja representada por apenas uma matriz, em que os valores de intensidade de $f(x, y)$ varie entre 0 e 1, então diz-se que esta é uma imagem binária, caso a variação seja maior, ela é dita ser uma imagem em escala de cinza. E se a representação ocorrer com mais de uma matriz, em que há diversos componentes representados, então

se tratará de uma imagem colorida. O que define como serão representadas as componentes de uma imagem são os espaços de cores [2, 16].

Espaços de Cores. Um espaço de cor é um padrão para especificação de cores como pontos em um sistema de coordenadas [2, 16]. Espaços de cores são necessários tanto para padronização da representação de cores, como para prover certas facilidades, em que a partir de um determinado espaço se pode obter mais facilmente algum resultado, como uma melhor limiarização de objetos [16]. Os dois espaços mais utilizados são explicitados a seguir.

RGB. A cor é o resultado da percepção da luz pelos olhos, sendo que a maioria das cores visíveis podem ser representadas pela combinação de luzes monocromáticas nos comprimentos de onda do vermelho, verde e azul [2, 16]. Imagens digitais coloridas são normalmente armazenadas nestas três componentes primárias, em que a partir das combinação destas se é possível obter outras. Este espaço de cor é o conhecido como RGB (do inglês, vermelho, verde e azul) [2, 16].

HSV. O HSV (do inglês, matiz, saturação e valor), também muito utilizado, é composto de três componentes [2, 16]: *Hue* ou Matiz: é a componente que define a cor propriamente dita. Pode variar de 0 a 360; *Saturation* ou Saturação: é usada para determinar qual a pureza da cor, varia em uma escala de 0 a 100; *Value* ou Valor: é o atributo utilizado para descrever o brilho, por exemplo, o preto possui valor 0. Também varia de 0 a 100.

A conversão entre espaços de cores também é possível e muito utilizada. Para se converter, por exemplo, de RGB para HSV, pode se prosseguir da seguinte forma: seja m o menor e M o maior valor entre as componentes R, G e B, e, normalizando tais componentes para que estejam entre 0 e 1, basta então aplicar as Equações 1, 2 e 3 para se encontrar, respectivamente, os valores das componentes H, S e V.

$$H = \begin{cases} 60 \times \frac{(G - B)}{(M - m)} + 0 & \text{se } M = R \text{ e } G \geq B \\ 60 \times \frac{(G - B)}{(M - m)} + 360 & \text{se } M = R \text{ e } G < B \\ 60 \times \frac{(G - B)}{(M - m)} + 120 & \text{se } M = G \\ 60 \times \frac{(G - B)}{(M - m)} + 240 & \text{se } M = B \end{cases} \quad (1)$$

$$S = \begin{cases} \frac{(M - m)}{(M)} & \text{se } M \neq 0 \\ 0 & \text{caso contrário} \end{cases} \quad (2)$$

$$V = M \quad (3)$$

Textura. Não há consenso sobre a melhor definição para texturas, pois em parte textura é algo percebido primariamente de forma sensorial por pessoas, mas podem ser descritas como conjuntos de padrões randômicos que aparecem regularmente em superfícies físicas, os quais quando analisados, podem ser utilizados para a identificação e descrição de determinada superfície [18, 2, 19].

Segmentação de Imagens. Uma etapa importante em processamento de imagens, é a separação de duas componentes de uma imagem: o fundo e objetos de interesse [2, 17]. A abordagem mais comum para a realização de segmentação é a utilização da técnica de *thresholding* ou limiarização, a qual é explicada a seguir.

Limiarização. A limiarização de uma imagem, ou de subespaços dela, consiste na definição de um limiar t para uma componente da imagem com o qual se consiga a melhor separação possível entre os objetos de interesse e o fundo desta [2, 17]. Seja o resultado esperado uma imagem binária, onde os objetos de interesse sejam representados pelo valor 1 (branco) e o fundo pelo valor 0 (preto), para cada ponto $f(x, y)$ da imagem aplica-se a Equação 4.

$$h(x, y) = \begin{cases} 1 & \text{se } f(x, y) > t \\ 0 & \text{se } f(x, y) < t \end{cases} \quad (4)$$

Background Subtraction. Esta técnica de segmentação consiste na obtenção de uma máscara através da subtração entre o fundo de uma imagem e outra onde há um objeto de interesse sobre esse mesmo fundo, de forma que, quando aplicado um algoritmo de limiarização se obtenha o contorno de tal objeto [2, 21]. Para sua realização, é necessário que haja no mínimo duas imagens: uma com apenas o fundo da cena e outra em que há o objeto de interesse sobre esta mesma cena, ou seja, ela somente se aplica em casos onde há um fundo estático durante a aquisição de imagens [2, 21].

Descrição e Extração de características. O primeiro passo antes de se iniciar o processo de classificação propriamente dito, é o de extrair características ou atributos que podem ser relevantes para a descrição de um determinado objeto ou características desejadas deste, como sendo pertencente a certa classe ou grupo [22, 23]. Para a representação de um objeto, um vetor de características é gerado, que é a representação dos atributos de interesse do objeto, o qual é resultado da junção das diversas características extraídas deste objeto [22, 23].

Descritores de cor. Considerando uma imagem colorida representada em um determinado espaço de cores, é possível descreve-la através de um vetor composto de características baseadas em medidas descritivas dos valores das componentes deste espaço de cores ao longo de toda imagem [2,16]. Entre as medidas mais comuns estão a de dispersão, que envolvem cálculos como média e desvio padrão, as quais foram utilizadas pelo método desenvolvido.

Descritores de forma. Dada uma forma, é comumente necessário quantificar algumas de suas característica, um processo chamado de caracterização de formas, em que isso pode ser obtido através de séries de medidas e propriedades extraídas da forma, formando assim um vetor próprio de características que descreve tal forma [17].

No presente trabalho, foi utilizada a descrição baseada na assinatura do contorno da forma, que, no presente caso, consiste das distâncias dos pontos da borda do objeto em relação ao seu centroide, sendo que em seguida é aplicada a Transformada Discreta de Fourier, obtendo-se assim um conjunto de características baseado nesta análise em frequência da assinatura [17].

Descritores de textura. Texturas podem variar de acordo com sua randomicidade, regularidade (ou periodicidade), direcionalidade e orientação [18]. Para sua descrição, as abordagens mais utilizadas são a estatística, espectral e estrutural [2]

A análise estatística consiste em analisar a textura de forma a descrevê-la por meio de medidas como o desvio padrão e a média, sendo portanto não possível a reconstrução de uma textura a partir destas características [2, 18]. A espectral consiste na análise em frequência, como na utilização da transformada de *Fourier*, descrevendo a periodicidade de forma global através da identificação de picos de energia no espectro de frequência [2, 18].

Já a abordagem estrutural, ocorre através da caracterização de texturas como sendo compostas de primitivas mais simples, em que estas primitivas seguem arranjos de forma regular [2, 18].

No presente trabalho foram utilizados alguns descritores de Haralick, onde há a descrição das texturas da imagem de uma forma global e estatística [24].

Reconhecimento de Padrões. Reconhecimento de padrões é a área cujo objetivo é a classificação de objetos em um número de classes, em que estas são um conjunto ordenado de categorias relacionadas usadas para agrupar dados de acordo com suas similaridades [22, 23]. Os conceitos que podem compreender um processo de reconhecimento de padrões, especialmente aqueles que foram utilizados por este trabalho, são: redução de dimensionalidade, classificadores e validação cruzada. Estes conceitos são brevemente explanados a seguir.

Redução de dimensionalidade. A redução de dimensionalidade ou seleção de características, consiste na análise dos vetores de descrição dos objetos, geralmente por meios estatísticos, resultando na eliminação de certas componentes, obtendo-se apenas aqueles atributos que mais contribuem para a descrição desses objetos [22, 23]. Essa redução é principalmente realizada com o objetivo de retirar ruídos, os quais são atributos que não acrescentam informação útil à descrição do objeto, o que também resulta na redução das necessidades computacionais dos classificadores durante a etapa de aprendizagem [22, 23].

Classificadores. Classificação é o processo pelo qual, dado um conjunto de vetores de características não categorizados em classes ou grupos, esses são analisados por um algoritmo classificador que os rotulam como pertinentes às classes que tal algoritmo julgar mais similar, seja com base em padrões aprendidos ou por análises de similaridade [22, 23]. Um classificador é, assim, um algoritmo que dadas várias entradas não conhecidas irá agrupá-las, por classe, através do reconhecimento de suas características em comum [22].

Devido ao problema de que não há métodos eficientes para se saber a priori qual o melhor classificador utilizar para determinado problema, a utilização de vários tipos se faz extremamente necessária, assim, dentre os classificadores mais comuns e que serão utilizados por este trabalho, podem-se destacar as árvores de decisão, as redes neurais artificiais e as *support vector machines*.

Validação cruzada. A divisão de um *dataset* entre diversos conjuntos de dados, alguns para treinamentos e outros para testes, se faz necessária para evitar resultados falsos, os quais seriam obtidos ao se testar os classificadores sobre a mesma base de dados sobre a qual eles construíram o seu modelo de classificação. Considerando-se que o estudo seja sobre um conjunto grande de dados o suficiente para ser dividido por n partes mais ou

menos iguais entre dados para treinamento e teste e, ainda gerar bons resultados, este processo resultaria em uma validação cruzada de n -vias [22, 23].

MATERIAIS E MÉTODOS

Materiais utilizados. Para a implementação do projeto, foi utilizada a linguagem de programação C++, através da IDE (*Integrated Development Framework* ou Ambiente de Desenvolvimento Integrado) QtCreator *Community* 3.4 com a utilização do *framework* Qt 5.5 para a construção da interface gráfica, tornando o projeto multi-plataforma [25].

O *hardware* utilizado tanto para o desenvolvimento como para processamento dos dados foi um computador do tipo *desktop*, com processador Intel Core i7 4930K de 3,40GHz, 16GB de memória RAM, HD de 1TB e monitor LCD de 23".

Framework WEKA. O WEKA (*Weikato Environment for Knowledge Analysis*) é uma coleção de algoritmos para aprendizado de máquina [26]. O *framework* possui um tipo de arquivo único para utilização, chamado de *arff*, o qual é constituído de um arquivo de texto ASCII para descrever um conjunto de atributos, os quais serão utilizados posteriormente por classificadores para a geração dos resultados finais.

Framework OpenCV. O OpenCV é um projeto *open source*, ou seja, possui código fonte livre, que contém diversas implementações de algoritmos para visão computacional. Entre as funções encontradas, há desde o carregamento de diferentes tipos de arquivos até algoritmos para as de conversão de imagens coloridas para a escala de cinza e espaço de cores HSV, o que será de extrema importância para o presente trabalho, além de funções mais complexas, como a implementações de algoritmos para extração de características [27]. Sua importância e relevância podem são comprovados pelos diversos trabalhos realizados utilizando esta biblioteca [27].

Banco de imagens. As imagens a serem utilizadas para o desenvolvimento do trabalho foram geradas por autoria própria, com o intuito de se criar um banco de imagens que posteriormente poderá ser disponibilizado publicamente. Este banco contém amostras representativas das 4 classes que serão estudadas por este trabalho: soja com danos por percevejo, mecânico e com esverdeado e soja sem danos. Uma imagem representativa desse conjunto é mostrada na Figura 5.

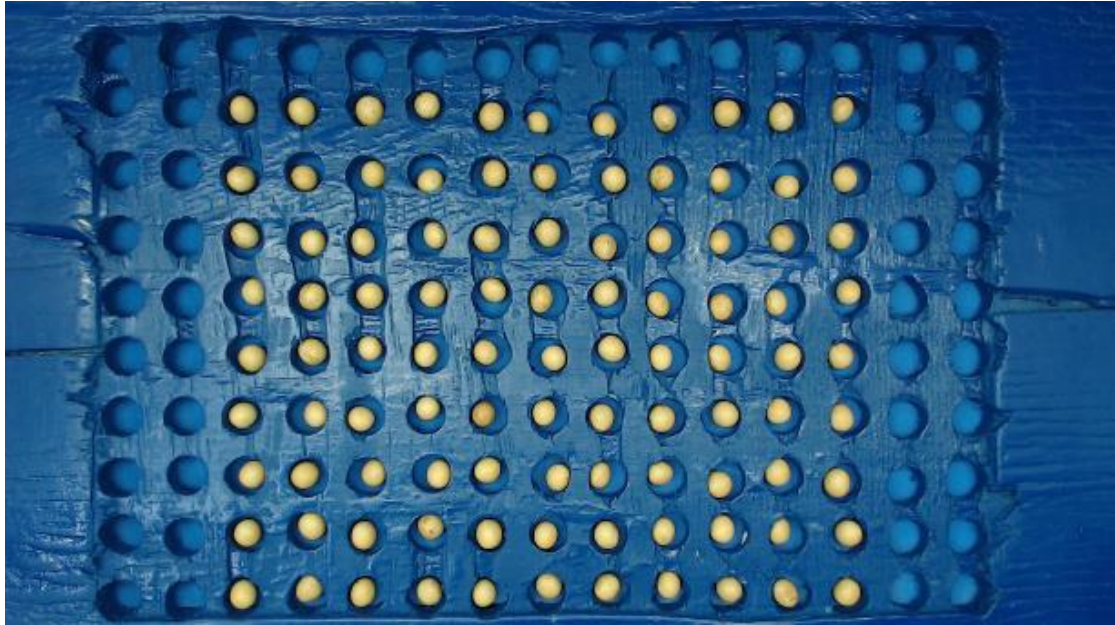


Figura 5. Imagem exemplo do banco de imagens, referente a classe de imagens saudáveis. Fonte: Autoria própria.

As imagens foram coletadas em 3 diferentes perspectivas com o intuito de se capturar toda a superfície de cada semente, sendo que as variações de perspectivas foram realizadas automaticamente por um equipamento projetado especificamente para esta coleta de imagens, sendo portanto, perspectivas com posicionamento não exato. Um modelo esquemático do equipamento pode ser visto na Figura 6.

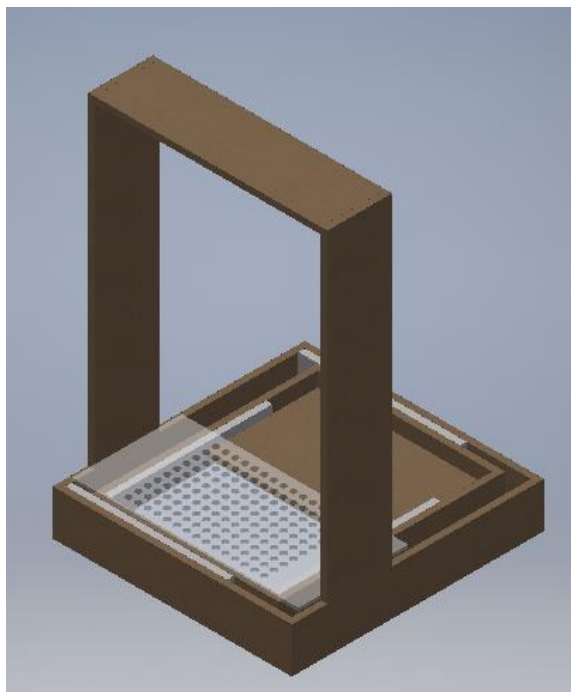


Figura 6. Figura representativa do equipamento de aquisição das imagens. Fonte: Autoria própria.

Este aparato de captura de imagens consiste de um Raspberry Pi 2, o qual é

responsável pelo controle de uma câmera webcam Logitech C920, 1 servo motor para a realização da movimentação das sementes e uma placa onde estas são colocadas.

Metodologia desenvolvida. Após a revisão da literatura e dos trabalhos relacionados, definiu-se a seguinte metodologia para o processamento das imagens e reconhecimento de padrões neste trabalho, dividida em 4 etapas:

1. Primeiramente, foi aplicada a técnica de segmentação de *Background Subtraction* com o objetivo de extrair especificamente as imagens das sementes de soja do restante do fundo da imagem.
2. Nesta etapa, ocorreu a extração de características através dos algoritmos e técnicas já citados na seção de Revisão Bibliográfica, tendo como resultado a geração dos vetores representativos das imagens, os quais serão guardados no formato de arquivo *arff*, próprio do *Weka*.
3. Foi então aplicada a técnica de redução de dimensionalidade nos vetores de descrição, utilizando o algoritmo *Correlation-based feature selection*, também do próprio *Weka*, cujos resultados foram utilizados na etapa final.
4. O arquivo gerado na etapa 3 foi então fornecido como entrada aos algoritmos classificadores do *Weka*, onde foi realizado o treinamento e classificação sobre todo o banco de imagens gerado através do método supervisionado, baseado em validação cruzada de 10-vias.

Um esquema representativo das etapas da metodologia é mostrado na Figura 7.

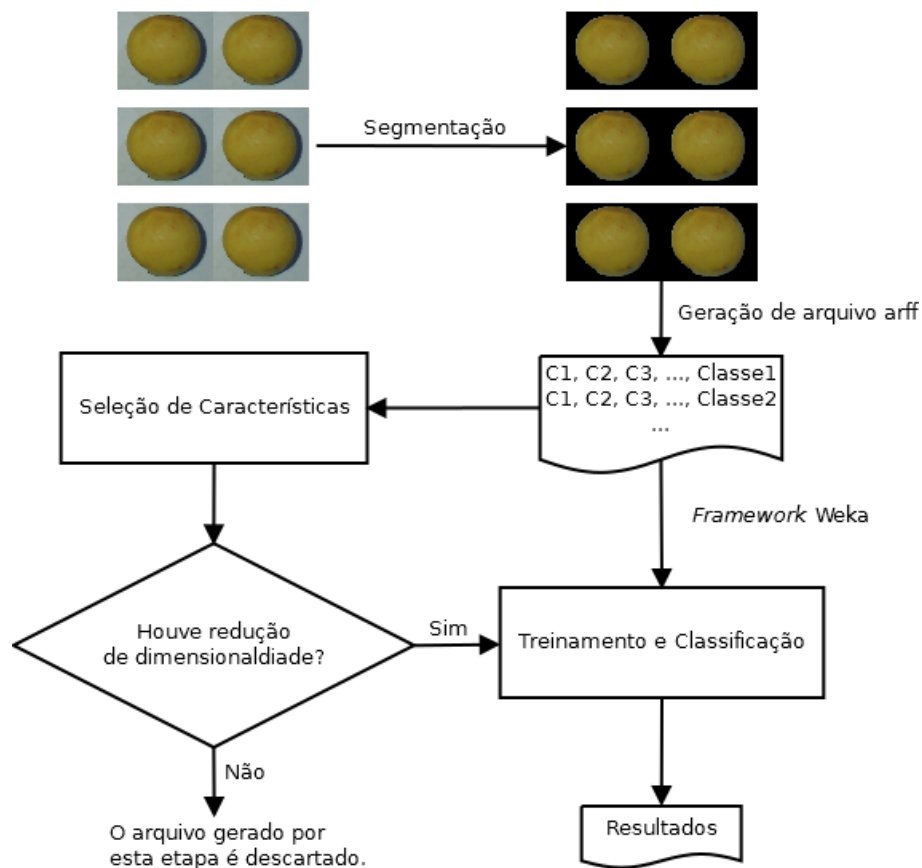


Figura 7. Esquema da metodologia proposta dividida em etapas. Fonte: Autoria própria.

RESULTADOS E DISCUSSÕES

Após o processamento do banco de imagens, o conjunto dos vetores de características resultante, após a realização do processo de redução da dimensionalidade, foi:

- 7 características de cor
- 100 características de forma
- 10 características de textura

Os resultados da classificação destes vetores pelos algoritmos do *Weka* estão apresentados na Tabela 1, onde há os valores de performance obtidos utilizando-se os classificadores IBK, J48, Naive Bayes (NB) e Random Forest (RF), MultiLayer Perceptron (MLP) e Support Vector Machine (SVM).

Tabela 1. Resultados referentes aos classificadores IBK, J48, NB, RF, SVM e MLP compreendendo 5 diferentes estatísticas de qualidade da classificação, considerando a média dos valores obtidos para todas as classes classificadas.

Classificador	Precisão	Recall	F-Score	Área ROC	Acurácia (%)
IBK	0,61	0,68	0,65	0,69	65,02
J48	0,66	0,69	0,71	0,76	72,10
NB	0,62	0,63	0,63	0,66	62,22
RF	0,70	0,70	0,71	0,75	70,31
SVM	0,78	0,77	0,78	0,80	78,73
MLP	0,80	0,84	0,83	0,91	81,25

A partir dos dados da Tabela 1, se verifica que o classificador do tipo rede neural MLP alcançou o melhor resultado geral, classificando corretamente 81,25% das imagens, seguido pelo SVM, com 78,73%, sendo que ambos apresentam o restante das estatísticas muito próximas da acurácia, indicando que os classificadores desempenharam de forma constante.

Enquanto que os classificadores baseados em árvores de decisão obtiveram os resultados intermediários, os piores, obtidos pelo IBK e NB, de 65,02% e 62,22% respectivamente, ajudam a concluir que para este tipo de problema, os classificadores de aprendizagem mais profunda são mais promissores.

CONCLUSÕES

O trabalho proposto teve como meta desenvolver um sistema capaz de analisar imagens de sementes de soja e classificá-las de acordo com a presença ou não de possíveis defeitos, além de descobrir as características e algoritmos que melhor se aplicam às deficiências que se desejam eliminar.

Com a utilização dos *frameworks* OpenCV e Weka, pôde-se obter resultados promissores, onde se alcançou 81,25% de acurácia com o melhor classificador, indicando que a metodologia desenvolvida, mesmo no estágio atual, se mostra muito confiável.

Através destes resultados, verifica-se a potencial eficácia que tal sistema poderia proporcionar no processo de beneficiamento de grãos de soja, necessitando apenas de mais alguns refinamentos para que se atinja níveis de qualidade e confiança no sistema suficientes para sua plena utilização em ambientes fabris.

Assim, espera-se que o desenvolvimento deste projeto, bem como seus resultados sejam de grande valia para avançar esta área de pesquisa, considerando-se principalmente que ainda há diversas possibilidades de melhoria para trabalhos futuros, como: implementar mais descritores de características para textura e forma; implementar um sistema visual para geração de relatórios de desempenho e ainda a possibilidade de se desenvolver um sistema completamente automático, o qual seria utilizado acoplado diretamente aos sistemas de beneficiamento de soja.

REFERÊNCIAS

- [1] FILHO, W. P. P. **Multimídia: conceitos e aplicações**. 2ª ed. Rio de Janeiro: LTC, 2011.
- [2] GONZALEZ, R. C. **Processamento de Imagens Digitais**. 3ª ed. São Paulo: Edgard Blucher, 2000.
- [3] WATANABI, S. **Pattern recognition: human and mechanical**. 1ª ed. Nova Iorque: John Wiley & Sons, Inc., 1985.
- [4] LIMA, G. V. L. d. et al. **Progress in pattern recognition, image analysis, computer vision, and applications: 20th iberoamerican congress, CIARP 2015**. In: Montevideo, Uruguay: Springer International Publishing, 2015. cap. **A Complex Network-Based Approach to the Analysis and Classification of Images**, p. 322–330.
- [5] BRILHADOR, A.; SERRARENS, D. A.; LOPES, F. M. **Progress in pattern recognition, image analysis, computer vision, and applications: 20th iberoamerican congress, CIARP 2015**. In: Montevideo, Uruguay: Springer International Publishing, 2015. cap. **A Computer Vision Approach for Automatic Measurement of the Inter-plant Spacing**, p. 219–227.
- [6] CEPEA. **Relatório PIBAgro-Brasil**. 2016. Disponível em: <http://www.cepea.esalq.usp.br/comunicacao/Cepea_PIB_BR_jan16.pdf>.
- [7] CONAB. **Acompanhamento da safra brasileira: grãos**. 2015. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/15_09_11_10_42_03_boletim_graos_setembro_2015.pdf>.
- [8] USDA. **World agricultural production**. 2016. Disponível em: <<http://apps.fas.usda.gov/psdonline/circulars/production.pdf>>.
- [9] BRASIL. **Regras para análise de sementes**. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 2009.
- [10] NETO, J. F.; KRZYZANOWSKI, F.; COSTA, N. d. **O teste de tetrazólio em sementes de soja**. Londrina: EMBRAPA-CNPSo, 1998.
- [11] FILHO, J. M.; KIKUTI, A. L. P.; LIMA, L. B. de. **Métodos para avaliação do vigor de sementes de soja, incluindo a análise computadorizada de imagens**. Revista Brasileira de Sementes, v. 31, n. 1, 2009.
- [12] OLIVEIRA, A.; SADER, R.; KRZYZANOWSKI, F. C. **Danos mecânicos ocorridos no beneficiamento de sementes de soja e suas relações com a qualidade fisiológica**. Revista Brasileira de Sementes, v. 21, n. 1, p. 59–66, 1999.
- [13] NETO, J. F.; HENNING, A. **Qualidades fisiológica e sanitária de sementes de soja**. Londrina, 1984.
- [14] PANIZZI, A. et al. **Efeitos dos danos de *piezodorus guildinii* (weestwood, 1837) no rendimento e qualidade da soja**. SEMINARIO NACIONAL DE PESQUISA DE SOJA, 1, 1978.
- [15] COSTA, N. P. et al. **Efeito de sementes verdes na qualidade fisiológica de sementes de soja**. Revista Brasileira de Sementes, v. 23, n. 2, p. 102–107, 2001.

- [16] VELHO, Z.; FRERY, A. C.; MIRANDA, J. *Image processing for computer graphics and vision*. 2a .ed. London: Springer, 2008.
- [17] CESAR, R. M.; Costa, L. F. *Shape classification and analysis: theory and practice*. 2ª ed. Flórida: CRC Press, 2009.
- [18] MIRMEHDI, M.; XIE, X.; SURI, J. *Handbook of texture analysis*. [S.l.]: Imperial College Press, 2008.
- [19] GOOL, L. V.; DEWAELE, P.; OOSTERLINCK, A. *Computer vision, graphics, and image processing*. In: [S.l.]: Elsevier, 1985. v. 29, n. 3, cap. Texture analysis anno 1983, p. 336–357.
- [20] PICCARDI, M. *Background subtraction techniques: a review*. In: IEEE. *IEEE international conference on systems, man and cybernetics*. [S.l.], 2004. v. 4, p. 3099–3104.
- [21] THEODORIDIS, S.; KOUTROUMBAS, K. *Pattern recognition*. 4th. ed. [S.l.]: Academic Press, 2008.
- [22] BISHOP, C. M. *Pattern recognition and machine learning*. Nova Iorque: Springer Science, 2006.
- [23] HARALICK, R. M. *Statistical and structural approaches to texture*. *Proceedings of the IEEE, IEEE*, 1979, 67, 786-804
- [24] QT. *Qt framework*. 2016. Disponível em: <<https://www.qt.io/qt-framework/>>.
- [25] HALL, M. et al. *The weka data mining software: an update*. SIGKDD Explorations, v. 11, n. 1, 2009.
- [26] BRADSKI, G.; KAEHLER, A. *Learning OpenCV: computer vision with the OpenCV library*. Cambridge, MA: O'reilly, 2008.