

Identificação e Classificação Automática de Sementes de Orquídeas

Adolfo M. Hengling¹, Edson T. Matsubara¹, Anderson V. de Araujo¹,
Eliton L. S. Perin¹, Mariane M. Hengling²,
Ceci C. Custodio², Nelson B. Machado-Neto²

¹Faculdade de Computação – Universidade Federal do Mato Grosso do Sul (UFMS)
Caixa Postal 549 – 79070-900 – Campo Grande – MS – Brasil

²Instituto de Biologia – UNOESTE
Pres. Prudente – SP – Brazil

{adolfohengling, elitonperin, mhengling}@gmail.com

{andvicoso, edsontm}@facom.ufms.br, {ceci, nelson}@unoeste.br

Abstract. *Counting the number of viable and non-viable seeds can be considered a laborious task that is often prohibitive in large experiments. An important part of the study of seeds is the accurate quantification of its germination. Recently computer vision and machine learning have been combined to design algorithms for similar tasks. In this article, we propose a solution to identify, classify and count orchid seeds automatically. The solution combines convex hull techniques along with nearest neighbour classification. The graphical, numerical and the significance test show that our proposal can deliver similar results to domain experts. We believe that the use of our proposal can enable large-scale experiments that are normally prohibitive using manual approaches because it takes only one tenth of the time needed by the manual approach to process the images.*

Resumo. *A contagem do número de sementes viáveis e não-viáveis pode ser considerada uma tarefa trabalhosa e que muitas vezes inviabiliza grandes experimentos. Uma parte importante para o estudo de sementes é a quantificação precisa da germinação dessas sementes em um experimento. Recentemente algoritmos que combinam aprendizado de máquina e visão computacional tem sido propostos para tarefas como esta. Neste artigo, é proposta uma solução que identifica, classifica e conta sementes de orquídeas automaticamente combinando uma técnica baseada em fecho convexo e vizinhos mais próximos. Segundo o teste de significância e também pelos resultados numéricos e gráficos, os resultados obtidos pelo sistema proposto tem desempenho similar ao desempenho de especialistas do domínio. A utilização da proposta pode assistir grandes experimentos em biologia que antes eram proibitivos pela contagem manual, pois necessita, aproximadamente, de apenas um décimo do tempo utilizado pela contagem manual para processar as imagens.*

1. Introdução

Orquídeas representam o maior número de espécies entre as Angiospermas [Chase et al. 2015]. Elas são geralmente utilizadas para decoração e algumas espécies

são importantes economicamente. A necessidade de criar bancos de armazenamento de sementes tem crescido desde que algumas espécies estão ameaçadas de extinção. Pesquisadores começaram a desenvolver estudos com o propósito de encontrar condições ideais para construir esses ambientes de preservação de sementes. Para testar um cenário experimental, é necessário avaliar indicadores tais como: taxa e velocidade de germinação e a vitalidade das sementes. O indicador vitalidade é obtido por meio da contagem do número de sementes viáveis e não-viáveis armazenadas na amostra que foi submetida ao cenário [Sawamura et al. 2014].

Apesar de ser considerada uma tarefa simples para humanos, contar e classificar sementes manualmente pode ser trabalhoso e tomar bastante tempo, principalmente em grandes experimentos. Essa tarefa precisa ser desempenhada cuidadosamente, pois é necessário determinar corretamente o sucesso ou a falha das condições adotadas a cada experimento. A condução de grandes experimentos é limitada pela dificuldade de processar grandes quantidades de imagens. O presente trabalho tem por objetivo reduzir a carga de trabalho manual dos pesquisadores pelo uso de uma ferramenta computacional que realize a classificação automática das sementes.

A literatura de biologia reporta poucos trabalhos que realizam a contagem automática de sementes viáveis e não viáveis, obtidas por coloração com solução do sal de tetrazólio, também conhecido como teste de tetrazólio. Assim, procurando em outros domínios como medicina foi encontrado um problema similar que conta células e que possuem o mesmo problema de sobreposição de sementes [Wienert et al. 2012]. Adaptando o modelo que corrige a sobreposição de células para o problema de sementes e adaptando o modelo de [Bieniecki and Grabowski 2004] para identificar as sementes, o presente trabalho desenvolve uma proposta original que utiliza diversos fundamentos de outros problemas similares.

Contribuições: A principal contribuição deste trabalho é o desenvolvimento da ferramenta computacional para identificação e contagem automática de sementes viáveis e não viáveis de orquídeas. As Figuras 1 e 2 ilustram um exemplo de resultado da aplicação da ferramenta proposta. A Figura 1 representa uma amostra de uma imagem real a ser processada pelo algoritmo, apresentando sementes viáveis e não viáveis. Um dos desafios da pesquisa deste trabalho é tratar a sobreposição das sementes e realizar a contagem corretamente. A Figura 2 ilustra o resultado obtido após o processamento da Figura 1 onde as sementes viáveis estão representadas com os contornos em vermelho e um ponto interno central preto e as sementes não viáveis com os contornos em branco e um ponto interno central verde.

A contagem automática de sementes torna o processo de obtenção dos indicadores requeridos mais fácil e rápido. Com isso, possibilita que os pesquisadores testem cenários alternativos com mais facilidade, pois a abordagem automática processa as imagens em apenas um décimo do tempo utilizado por um especialista do domínio. Nos experimentos, não houve diferença significativa entre a contagem manual realizada por um especialista e a contagem realizada pela proposta.

2. Trabalhos Relacionados

Existem duas grandes fases distintas neste trabalho: (1) segmentação da imagem e (2) identificação e separação de objetos sobrepostos.

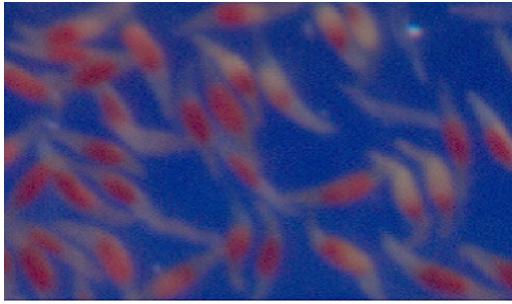


Figura 1. Imagem anterior ao processamento.

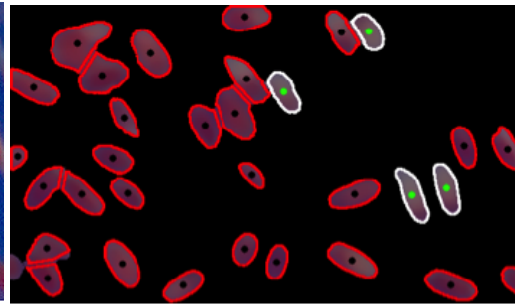


Figura 2. Imagem após o processamento.

A principal técnica aplicada na segmentação da imagem foi introduzida em [Bieniecki and Grabowski 2004]. Essa abordagem utiliza a informação de cor de cada pixel da imagem que é transformado em um vetor de características, e posteriormente é utilizado para treinar um classificador *K-Nearest Neighbors* (KNN). Após o treinamento, o classificador prediz a classe de cada pixel da imagem.

Devido ao alto número de sementes em uma amostra, algumas delas podem ficar sobrepostas umas as outras. Neste artigo, foi adaptado o algoritmo designado para a separação de células sobrepostas, descrito em [Wienert et al. 2012], ao problema da separação das sementes.

A segmentação também é utilizada para encontrar núcleo de células no trabalho de [Al-Kofahi et al. 2010], além de produzir um algoritmo robusto e acurado para realizar a segmentação. O algoritmo foi incluído no *toolkit* FARSIGHT. O projeto FARSIGHT desenvolve um conjunto de ferramentas de análise de imagem de microscópios ópticos modernos.

Os trabalhos descritos em [Joosen et al. 2010, Skrubek et al. 2015, Lurstwut and Pornpanomchai 2016] apresentam objetivos similares aos objetivos deste estudo, pois processam imagens a fim de encontrar e classificar sementes. Os autores apresentaram um estudo detalhado para a quantificação da germinação de sementes de várias espécies de plantas, tomate e arroz. Porém, estas abordagens limitam o número de sementes presentes nas amostras. As sementes são estrategicamente alocadas para evitar o problema da sobreposição de objetos. Portanto, isto enfatiza a importância deste trabalho, e também a necessidade de pesquisa neste campo.

3. Adaptação do algoritmo de separação de células

A ideia utilizada para a separação de sementes sobrepostas foi baseada em um algoritmo designado para a identificação e quantificação de células em imagens obtidas por meio de microscópios, descrito por [Wienert et al. 2012]. Neste trabalho, a solução proposta pelos autores é composta por seis passos, começando pela identificação de contornos fechados até a classificação dos objetos encontrados em núcleo de células ou outros objetos. Porém, para este trabalho foi considerado apenas o passo que descreve o procedimento para separação de objetos sobrepostos pela análise de concavidade. Este procedimento consiste em:

1. Encontrar concavidades no contorno das células: as concavidades representam anomalias no formato de uma célula. Esta característica é um forte indício de que o contorno representa uma sobreposição de várias células. Assim como as células, as sementes de orquídea também seguem um padrão em relação à seu formato. A Figura 3 é um exemplo de sementes individuais. As sementes individuais podem ser facilmente detectadas e classificadas pelo algoritmo desenvolvido. A Figura 4 é um exemplo de dois agrupamentos de sementes. É possível observar que podemos utilizar a mesma técnica e explorar as concavidades dos contornos para separar as sementes. Os detalhes de como identificar as concavidades de sobreposição de sementes estão descritos na Seção 4, Item 3a.



Figura 3. Sementes individuais.



Figura 4. Sementes sobrepostas.

2. Encontrar o melhor ponto de separação: após identificar as concavidades de um contorno, todos os pontos de todas as concavidades são combinados. De acordo com [Wienert et al. 2012], a linha de corte mais adequada (PontoInicial, PontoFinal) é formada pela combinação dos pontos que obtiveram a maior pontuação dada pela equação *SeparationScore* (1). Segundo os autores, esta equação previne que cortes sejam realizados em direções errôneas por meio do balanceamento efetuado pela equação *AngleScore* (3). De acordo com a Figura 5 (a), C_1 é o ponto inicial e C_2 o ponto final de uma potencial linha para corte (C_1, C_2) de comprimento r . A_1 e B_1 são os pontos extremos da concavidade 1, assim como A_2 e B_2 são os pontos extremos da concavidade 2. $depth_1$ é a distância de C_1 para o segmento formado por (A_1, B_1) e $depth_2$ é a distância de C_2 para o segmento formado por (A_2, B_2). α_1 é o ângulo entre C_1C_2 e (A_1, B_1). Da mesma forma, α_2 é o ângulo entre C_2C_1 e (A_2, B_2).

$$LengthScore = \frac{r}{r + depth_1 + depth_2} \quad (1)$$

$$AngleScore = \frac{\alpha_1 + \alpha_2}{2\pi} \quad (2)$$

$$SeparationScore = LengthScore + AngleScore \quad (3)$$

3. Separação de objetos: o processo descrito nos itens anteriores é executado recursivamente até que o contorno alcance um tamanho mínimo de referência. O tamanho mínimo (*avgSeedLength*) é definido pela média do perímetro do contorno de todas as sementes individuais da imagem (não sobrepostas). A condição de parada deste método é alcançada quando o perímetro do contorno analisado é menor que $avgSeedLength * \gamma$ - Seção 5.1. Ao final deste processo, assim como demonstrado na Figura 5 (b), assume-se que cada contorno na imagem representa uma semente individual.

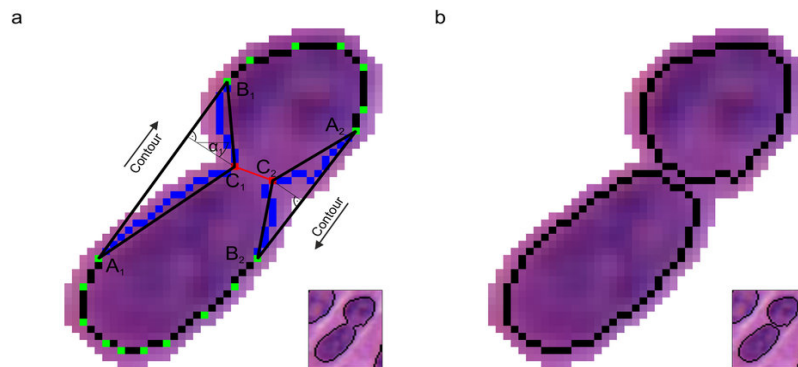


Figura 5. Exemplificação do método de separação de objetos [Wienert et al. 2012].

4. Proposta

Visando produzir uma ferramenta que solucione tanto o problema de encontrar e classificar as sementes de uma imagem e que também seja capaz de separar as sementes sobrepostas, o objetivo foi dividido em problemas menores. A Figura 6 ilustra os passos do processo adotado por esta solução.

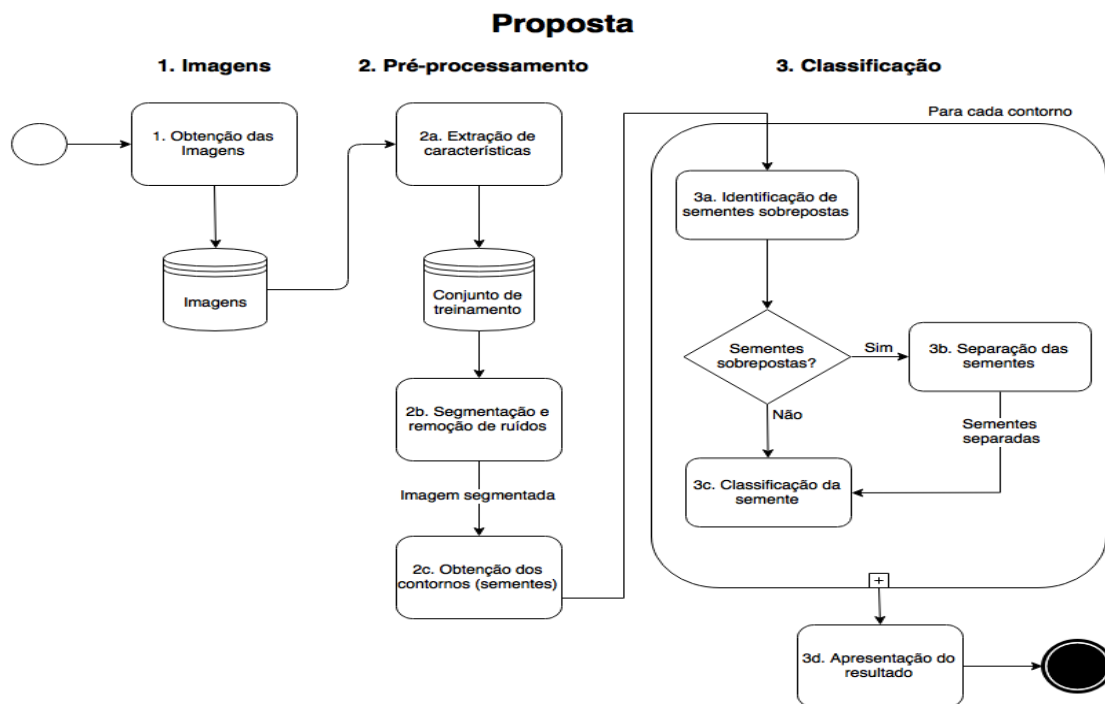


Figura 6. Passos do processo adotado.

1. Obtenção das imagens: as sementes foram armazenadas por zero, três, seis e nove meses. Foram utilizados 2 mg de sementes em três repetições. As sementes foram condicionadas com solução de sacarose 10% por 24 horas em temperatura ambiente. Após foram lavadas com água deionizada e incubadas com 1 ml de TZ1% por 24 horas no escuro a 40°C [Hosomi et al. 2011]. Após período de incubação parte da solução é drenada, as sementes são depositadas em lâminas

de microscopia e as imagens são capturadas em escâner de mesa HP G2710, com resolução de 3600 dpi. Para melhor contraste um fundo azul é adicionado à tampa do escâner.

2. Pré-processamento: no pré-processamento a entrada são as imagens das sementes e a saída são os contornos que delimitam cada semente ou grupo de sementes, no caso de sementes sobrepostas.

- (a) Extração de características: para obter o conjunto de treinamento, foi desenvolvida uma aplicação que armazena os valores RGB e a classe de um determinado pixel no formato [R,G,B,Classe] em um arquivo texto. Por meio desta aplicação é possível anotar a imagem como demonstrado na Figura 7. Nesta imagem foram anotadas quatro classes: sementes viáveis - representadas pela cor azul, sementes inviáveis - representadas pela cor vermelha, a palha das sementes - representada pela cor preta e o fundo da imagem - representado pela cor branca. O arquivo texto exportado é lido pela aplicação que processa a imagem. Estas características são utilizadas para treinar o classificador.

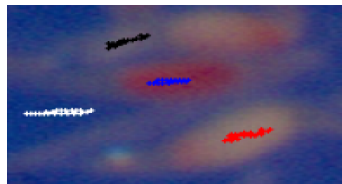


Figura 7. Extração de características de uma imagem.

- (b) Segmentação e remoção de ruídos: o objetivo desta etapa é separar a imagem em duas regiões: *background* e *foreground*. Para isso, o conjunto de características da imagem que foi extraído anteriormente é utilizado no treinamento para um classificador KNN [Cover and Hart 1967]. O classificador atribui uma classe a cada pixel da imagem. Os pixels que não pertencem às classes viável ou inviável são definidos como *background* e assumem a cor preta (RGB[0,0,0]). Caso contrário, o pixel permanece com sua cor original. Após a segmentação, um filtro para remoção de ruídos é aplicado para corrigir pequenas falhas do KNN. As Figuras 8 e 9 ilustram o resultado deste processo.
- (c) Obtenção dos contornos: os contornos são obtidos por meio da execução do método `findContours`, disponibilizado pela biblioteca OPENCV [Bradski 2000]. Um contorno é uma lista de pontos (x,y) que representa uma ou mais sementes, em caso de agrupamento. Os contornos identificados nesta fase servem como entrada para a próxima etapa do processamento. Os contornos com perímetro menor que θ são considerados ruído, e são descartados - Seção 5.1. A Figura 10 ilustra o resultado desta etapa. É possível identificar os agrupamentos entre as sementes que

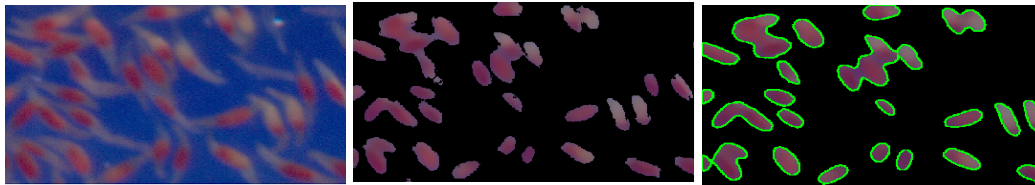


Figura 8. Imagem inicial.

Figura 9. Imagem segmentada.

Figura 10. Contornos obtidos.

precisam ser recortados.

3. Classificação:

- (a) Identificação de sementes sobrepostas: o agrupamento de sementes - ou sobreposição de sementes - pode ser identificado por meio do reconhecimento de concavidades no contorno do agrupamento. A análise para encontrar concavidades em um agrupamento de sementes é realizada através da análise do ângulo formado entre cada pixel do contorno e seus n vizinhos. Por causa da irregularidade dos contornos devido à baixa qualidade das imagens, foi necessário a aplicação de um algoritmo para obter descritores elípticos do contorno. Desta forma, um novo contorno é gerado com um número menor de elementos, de acordo com o parâmetro ϕ - Seção 5.1. Assim, é possível diminuir o número de falsas concavidades encontradas. A Figura 11 é um exemplo de contorno sem suavização. As irregularidades deste contorno são notáveis e induzem o algoritmo a considerar falsas concavidades. A Figura 12 ilustra o contorno da Figura 11 após suavização. As concavidades encontradas neste contorno estão destacadas em vermelho.

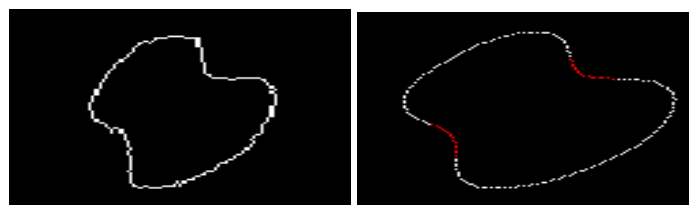
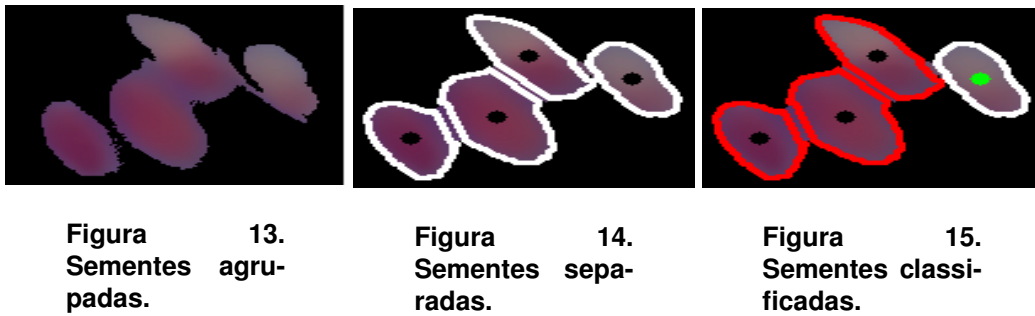


Figura 11. Contorno original.

Figura 12. Contorno suavizado.

- (b) Separação de sementes sobrepostas: após identificar os agrupamentos de sementes, o algoritmo adaptado para separação de sementes descrito neste artigo é executado - Seção 3. A Figura 13 ilustra um agrupamento de sementes antes do procedimento utilizado para separação. O resultado desta etapa pode ser observado na Figura 14. As sementes da Figura 13 foram separadas e estão representadas com contornos em branco e um ponto interno preto. Com isso, assume-se que todos os contornos representam sementes individuais e podem ser classificados como sementes viáveis ou não-viáveis.



(c) Classificação das sementes: nesta fase, é esperado que todos os contornos representem sementes individuais. Dessa forma, todos os pixels pertencentes à região interna de um determinado contorno são analisados pelo classificador KNN. Um pixel é classificado de acordo com a classe de seus κ vizinhos mais próximos - Seção 5.1. A semente é classificada de acordo com a classe dominante de pixels dentro de seu contorno. Isto é, se o número de pixels de classe viável é maior do que o número de pixels de classe inviável, então a semente é considerada viável, e vice-versa. A Figura 15 exemplifica o resultado desta etapa, onde as sementes viáveis estão representadas com os contornos em vermelho e um ponto interno central preto e as sementes não-viáveis como os contornos em branco e um ponto interno central verde.

(d) Apresentação dos resultados: após a classificação de todas as sementes encontradas, o resultado é informado ao usuário. Os dados informados são: total de sementes encontradas, número de sementes viáveis e inviáveis. Com isso, os pesquisadores são capazes de obter os indicadores necessários para avaliar o sucesso ou falha do experimento.

O aprendizado de máquina neste trabalho consiste em encontrar os diversos parâmetros do modelo que separa as sementes sobrepostas, do identificador e contador de sementes. O ajuste destes parâmetros é realizado utilizando *RandomSearch* [Bergstra and Bengio 2012]. A fase de treinamento consiste em treinar o KNN interno que faz a segmentação e a identificação das sementes, ajustar os modelos juntamente com os parâmetros descritos na Tabela 1 e a fase de teste passa por todos os modelos aprendidos que ao final conta as sementes em viáveis e não viáveis.

5. Experimentos

5.1. Conjunto de dados e algoritmos utilizados

O conjunto de 40 imagens utilizadas no desenvolvimento deste projeto foram produzidas por um estudo que visava armazenar sementes de orquídeas em diversas temperaturas pelo maior período de tempo possível, evitando assim a perda de viabilidade das mesmas, uma vez que o armazenamento pode causar estresse ou injúrias aos tecidos das sementes. As sementes foram depositadas em lâminas de microscopia, e imagens foram adquiridas por meio da digitalização dessas lâminas - Seção 4, Item 1. Em seguida, três regiões de cada imagem foram escolhidas para serem quantificadas. O processo de quantificação manual

Tabela 1. Ajuste de parâmetros utilizando *Randomized Search* utilizando 10 iterações com validação cruzada interna de três partições.

Simb.	Intervalo	Melhor parâmetro - Viável			Melhor parâmetro - Inviável		
		Part. 1	Part. 2	Part. 3	Part. 1	Part. 2	Part. 3
κ	[1..20]	5	17	6	7	18	3
γ	[1,5..2]	1,5	2	1,5	1,6	2	1,8
θ	[10..50]	35	15	20	20	20	40
ϕ	[5..20]	5	5	6	7	15	6

leva cerca de 15 minutos por imagem, enquanto o processo automático proposto por este estudo leva, em média, 1min49s por imagem.

A ferramenta de contagem foi desenvolvida utilizando a linguagem PYTHON versão 2.7.6, OPENCV versão 3.0 e KNN do SKLEARN versão 0.17.1. A implementação do algoritmo KNN da implementação do SKLEARN faz otimizações no espaço de exemplos com `BallSearch` ou `KDTree` [Goldberger et al. 2004] para redução de tempo de classificação.

Como mencionado nas seções anteriores, existem diversos parâmetros que precisam ser ajustados. [Bergstra and Bengio 2012] propõe um método para ajuste de parâmetros com resultados melhores que a popular `GridSearch`. Assim, as implementações usadas neste trabalho foram adaptadas para funcionarem com o `sklearn.grid_search.RandomizedSearchCV` para encontrar os melhores parâmetros da proposta. Foram utilizadas 10 iterações do `RandomizedSearchCV`, o que representa uma busca dos melhores 10 parâmetros por meio de 10 execuções da proposta. A Tabela 1 ilustra o intervalo de valores testados com o `RandomizedSearchCV` e o melhor parâmetro encontrado. Para evitar *overfitting*, é executada uma validação cruzada interna de três partições para definir os melhores parâmetros.

As métricas para avaliar os resultados foram:

$$MeanAbsoluteError(MAE) = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (4)$$

$$MeanAbsolutePercentageError(MAPE) = \frac{1}{n} \sum_{i=1}^n \left| \frac{f_i - y_i}{f_i} \right| \quad (5)$$

Onde n representa o número de imagens, f_i o número de sementes obtidas pela contagem manual e y_i o número de sementes obtidas pela contagem automática realizada pela proposta. No entanto essas métricas podem esconder erros de predição privilegiando a contagem para mais ou para menos, pois as métricas simplesmente somam os valores absolutos destas diferenças.

Para uma análise mais completa foi utilizado *Receiver Operating Characteristic for Regression Models* (RROC), introduzida por [Hernández-Orallo 2013]. A RROC estende uma das técnicas mais populares para análise gráfica do desempenho de classificadores que é a gráfico ROC [Fawcett 2006]. A técnica RROC consiste separar

a contagem de erro $e_i = y_i - f_i$ em dois conjuntos: $OVER = \sum_i \{e_i | e_i > 0\}$ e $UNDER = \sum_i \{e_i | e_i < 0\}$. O $OVER$ representa a soma dos erros de quando $y_i > f_i$, ou seja, a contagem das sementes da proposta é maior que a contagem manual. O $UNDER$ representa a situação inversa, a contagem das sementes da proposta é menor que a contagem manual.

5.2. Resultados

A avaliação experimental tem como objetivo verificar se a contagem de sementes viáveis e não viáveis pela contagem automática da proposta é equivalente a contagem manual humana feita por especialistas. Todos os resultados desta seção foram obtidos pela validação cruzada de três partições. A Figura 16 ilustra essa comparação na qual nos eixos X e Y representam respectivamente o número de sementes contadas pela proposta e pela contagem manual. A contagem idêntica entre a contagem da proposta e da comparação manual acontece quando os pontos são representados em cima da linha diagonal azul.

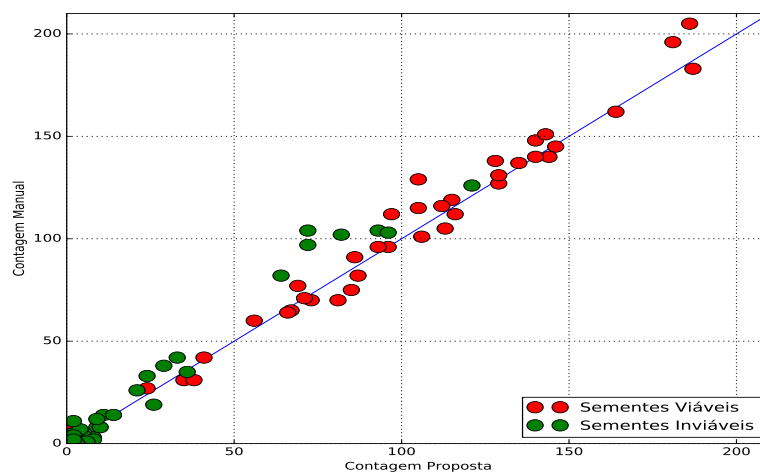


Figura 16. Contagem manual e Contagem da proposta.

Pela figura, tanto a contagem para sementes viáveis quanto para sementes inviáveis ficam bastante próximas da contagem do especialista. Uma observação interessante a ser ressaltada está na pequena variância da distância entre diagonal e os pontos, mesmo nos pontos mais afastados da origem. Em outras palavras, isso indica que o erro da proposta permanece quase constante mesmo para imagens com maior número de sementes.

O $MAPE$ do sistema foi de 9,19% com desvio padrão de 9,54. Assim, em uma imagem com 100 sementes o sistema erra aproximadamente 9,19, contanto as sementes para mais ou para menos. Uma outra métrica bastante utilizada é a MAE que foi de 10,05 com 9,8 de desvio padrão com imagens com 119,75 sementes por imagem na média. Dessa forma, para cada imagem espera-se que a contagem esteja errada em 10,05 sementes, novamente contando para mais ou para menos.

Para melhor distinguir a contagem para mais ou para menos, foi utilizado o RROC. As Figuras 17 e 18 representam as RROC para as viáveis e para as inviáveis. A contagem

perfeita, sem nenhum erro, seria apresentada no gráfico na posição (0,0) no canto superior esquerdo. Portanto, quanto mais próximo de (0,0) melhor é o sistema. Neste caso, ambas as duas contagens estão bem próximas do ideal.

Um ponto que está acima da diagonal que parte do canto superior esquerdo para o canto inferior direito representa uma contagem maior de *OVER* do que de *UNDER*. No caso dos experimentos deste trabalho, pontos acima da diagonal representam uma contagem maior pelo sistema do que pela contagem manual. Em outras palavras, o sistema, quando erra na contagem das sementes, normalmente tem a tendência de errar para mais do que para menos. É possível observar que esta situação ocorre tanto para sementes viáveis quanto para sementes inviáveis.

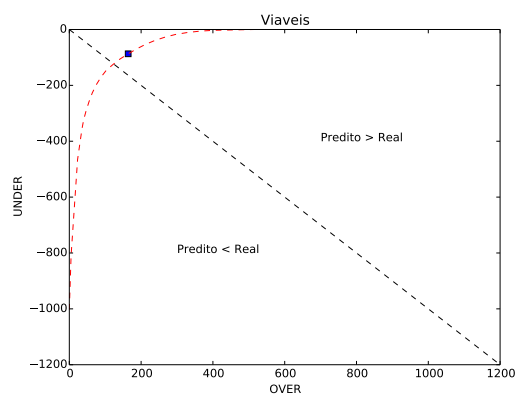


Figura 17. RROC viável

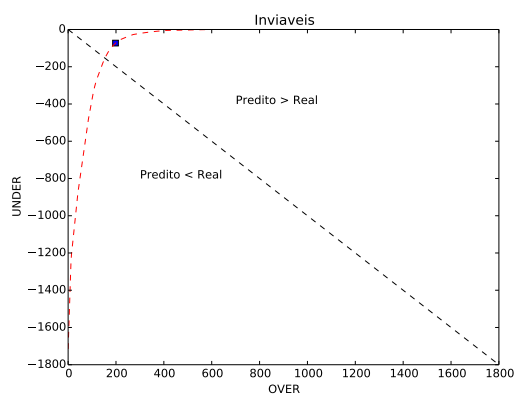


Figura 18. RROC inviável

Como pode ser observado o número encontrado na contagem manual e na contagem realizada pela abordagem proposta são muito similares. Para verificar se existe diferença significativa entre as contagens foi realizado um teste t não pareado bicaudal com 95% de grau de confiança. Considerando médias de 119,75 e 31,28 de desvio padrão para a contagem manual e 124,80 e 32,85 de desvio padrão para contagem da proposta com uma amostra de 40 imagens obteve-se um intervalo de confiança entre [-19.33,9.23] e a diferença das médias foi de -5.05. Portanto não se pode rejeitar a hipótese nula de que as amostras são similares. Desse modo, pode-se afirmar que para o conjunto de imagens testados, não existe diferença significativa entre a contagem manual e a contagem realizada pela proposta.

6. Conclusão

Neste trabalho é proposta uma ferramenta que combina algoritmos de classificação e processamento de imagens para contar sementes viáveis e não viáveis de orquídeas obtidas de imagens de microscópios. Trabalhos anteriores apresentam diversos componentes deste trabalho que possibilitam a segmentação, identificação, resolução de células sobrepostas e a contagem. O presente trabalho propõem uma combinação bem sucedida dessas propostas.

Os resultados são promissores. Gráficos, métricas e o teste de significância indicam que a contagem manual humana realizada por especialistas do domínio e pela ferramenta proposta são similares.

Como trabalhos futuros, pretende-se colocar a ferramenta desenvolvida dentro de um servidor de internet para que possa ser acessada publicamente pela comunidade científica. Além disso, pretende-se estudar e comparar o resultado obtido por este trabalho com outros classificadores, como *redes neurais*, *árvores de decisão* e *SVM*, e também com outras formas de realizar a separação de sementes utilizando, por exemplo, o algoritmo *watershed* e outras técnicas que possuem similaridade com a proposta.

Referências

- Al-Kofahi, Y., Lassoued, W., Lee, W., and Roysam, B. (2010). Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering*, 57(4):841–852.
- Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305.
- Bieniecki, W. and Grabowski, S. (2004). Nearest neighbor classifiers for color image segmentation. *Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2004. Proceedings of the International Conference*, (8230798):209 – 212.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Chase, M. W., Cameron, K. M., Freudenstein, J. V., Pridgeon, A. M., Salazar, G., Berg, C. V. D., and Schuiteman, A. (2015). An updated classification of orchidaceae. *Botanical Journal of the Linnean Society*, 177(2):151–174.
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recogn. Lett.*, 27(8):861–874.
- Goldberger, J., Hinton, G. E., Roweis, S. T., and Salakhutdinov, R. (2004). Neighbourhood components analysis. In *Advances in neural information processing systems*, pages 513–520.
- Hernández-Orallo, J. (2013). Roc curves for regression. *Pattern Recognition*, 46(12):3395–3411.
- Hosomi, S. T., Santos, R. B., Custodio, C. C., Seaton, P. T., Marks, T. R., and Machado-Neto, N. B. (2011). Pre-conditioning cattleya seeds to improve the efficacy of the tetrazolium test for viability. *Seed Sci Technol.*, 139:178–189.
- Joosen, R. V. L., Kodde, J., Willems, L. A. J., Ligterink, W., van der Plas, L. H. W., and Hilhorst, H. W. (2010). germinator: a software package for high-throughput scoring and curve fitting of arabidopsis seed germination. *The Plant Journal*, 62(1):148–159.
- Lurstwut, B. and Pornpanomchai, C. (2016). Application of image processing and computer vision on rice seed germination analysis. *International Journal of Applied Engineering Research*, 11(9). ISSN 0973-4562.
- Sawamura, L. H., Neto, N. B. M., and Hosomi, S. T. (2014). Conservação de sementes prégerminadas de orquídeas tropicais em criocongelamento. *Colloquium Agrariae*, 10(Especial):31–37. ISSN: 1809-8215. DOI: 10.5747/ca.2014.v10.nesp.000134.
- Skrubej, U., Rozman, C., and Stajniko, D. (2015). Assessment of germination rate of the tomato seeds using image processing and machine learning. *European Journal of Horticultural Science*, pages 68–75. DOI: 10.17660/eJHS.2015/80.2.4.
- Wienert, S., Heim, D., Saeger, K., Stenzinger, A., Beil, M., Hufnagl, P., Dietel, M., Denkert, C., and Klauschen, F. (2012). Detection and segmentation of cell nuclei in virtual microscopy images: A minimum-model approach. *Scientific Reports*, 2(503).