

Detecção automática de bagas de café em imagens de campo

Thiago T. Santos¹

¹Embrapa Informática Agropecuária
Caixa Postal 6.041 – 13083-886 – Campinas – SP – Brasil
thiago.santos@embrapa.br

RESUMO

Um método para detecção automática de bagas de café em imagens digitais tomadas em campo é apresentado. Ele consiste em uma fase de detecção preliminar de regiões candidatas seguida por uma etapa de classificação capaz de filtrar falsos positivos. Para o treinamento e avaliação da classificação automática, 3.393 imagens foram manualmente rotuladas. Testes quantitativos demonstraram a identificação automática de bagas de café com 90% de precisão através de Máquinas de Vetores de Suporte e descritores HOG (Histograma de Gradientes Orientados). Testes qualitativos indicam que o classificador treinado exibe bons resultados em novas imagens de campo. Esses resultados evidenciam que a detecção automática de frutos é factível e pode ser aplicada em metodologias de predição de safra e sistemas de agricultura de precisão.

PALAVRAS-CHAVE: Aprendizado de máquina, Visão Computacional, Reconhecimento, Fruticultura.

ABSTRACT

A method for automatic detection of coffee berries in digital images acquired in the field is presented. It consists of a preliminary detection step that finds candidate regions, followed by a classification stage which filters out false positives. For training and evaluation of automatic classification, 3,393 images were manually labeled. Quantitative tests have shown the automatic identification of coffee berries with 90% accuracy using Support Vector Machines and HOG descriptors (Histogram of Oriented Gradients). Qualitative tests indicate that the trained classifier exhibits good results for new field images. These results show that automated berry detection is feasible and can be applied in crop prediction methodologies and precision farming systems.

KEYWORDS: Machine Learning, Computer vision, Recognition, Fruit production.

INTRODUÇÃO

A detecção automática de frutos no campo é parte essencial de aplicações em predição de safra, agricultura de precisão e fenotipagem para melhoramento vegetal, além de ser uma etapa fundamental na automação de atividades como colheita e aplicação de insumos (LEE et al., 2010). Devido ao tamanho dos frutos, o sensoriamento remoto é inviável, de forma que o *sensoriamento proximal* realizado por sensores posicionados entre as linhas de produção torna-se o mais indicado. Além de não ser invasivo, o imageamento pode ser realizado por câmeras movimentadas por operadores humanos ou acopladas a tratores, *rovers* ou drones.

A coleta de dados fenológicos em campo para predição de safra ou fenotipagem, se conduzida por operadores humanos, constitui uma atividade lenta, laboriosa e imprecisa. Por outro lado, a detecção automática de frutos por sensoriamento proximal é capaz de aumentar o número de amostras obtidas em campo, ao mesmo tempo em que minimiza a variação das medições, permitindo assim a aquisição de volumes maiores de dados, necessários tanto à fenotipagem em larga escala (FURBANK; TESTER, 2011) quanto à agricultura de precisão baseada em *Big Data* (SONKA, 2014).

Nuske et al. (2011) empregaram a Transformada de Simetria Radial (LOY; ZELINSKY, 2003) para detectar regiões em imagens de vinhas que poderiam corresponder a bagas de uva. De cada região foi extraído um descritor, composto por valores referentes a dados de cor e textura obtidos na imagem. Os autores selecionaram um conjunto de descritores e realizaram sua classificação manualmente, produzindo assim um conjunto de treinamento para uma etapa de classificação supervisionada. O algoritmo de K -vizinhos (HASTIE; TIBSHIRANI; FRIEDMAN, 2009) foi então utilizado na classificação dos descritores, identificando regiões como bagas. Em um refinamento final, os autores eliminaram regiões classificadas como bagas, mas que ocorriam isoladamente, pressupondo que os frutos deveriam aparecer em grupos correspondentes aos cachos de uvas.

A mesma estratégia, detecção de formas circulares, filtragem por um classificador estatístico e aglomeração de bagas em cachos, também é empregada por Roscher et al. (2014). Seu sistema utilizou a Transformada Circular de Hough (YUEN et al., 1990) para identificar regiões candidatas na imagem. Cada região foi então representada por descritores HOG (DALAL; TRIGGS, 2005) e *gist* (OLIVA; TORRALBA, 2006). Finalmente, um classificador baseado em campos aleatórios condicionais (LAFFERTY; MCCALLUM; PEREIRA, 2001) foi utilizado na classificação de cada candidato nas classes *baga* e *não baga*. Uma vantagem no uso dos campos condicionais é que, no processo de otimização do classificador, há uma penalização para candidatos espacialmente vizinhos atribuídos a classes diferentes, o que favorece a identificação de cachos e o descarte de candidatos que ocorram isoladamente na imagem.

De acordo com nosso conhecimento até o presente momento, não há estudo similar envolvendo a detecção automática de frutos de café em imagens de campo. Um sistema de detecção automática de bagas de café teria ao menos duas aplicações imediatas. A primeira é a estimação de safra a partir de medidas fenológicas. A estimativa de safra de café é um problema impor-

tante, pois impacta a cotação da *commodity* nas Bolsas de Valores (MIRANDA; REINATO; SILVA, 2014). Alguns modelos matemáticos foram propostos para prever a produtividade de talhões de café a partir de atributos fenológicos coletados manualmente, como a altura e o diâmetro da base da planta e o número de frutos observados no 4.º e 5.º internódios a partir do ápice do ramo plagiotrópico (FAHL et al., 2005; OLIVEIRA, 2007; MIRANDA; REINATO; SILVA, 2014). Os frutos são contados manualmente em poucos internódios e as dimensões da planta são utilizadas para extrapolar o possível número de frutos no restante do cafeeiro. A coleta manual de dados se limita a poucas plantas no talhão: Miranda, Reinato e Silva (2014), por exemplo, utilizaram 10 plantas por talhão. Sendo assim, a detecção automática de bagas permitiria uma amostragem maior tanto dos frutos quanto das plantas, além de evitar erros e inconsistências comuns em anotação manual. A segunda aplicação seria a avaliação da maturação de frutos. Os frutos de café em um mesmo ramo podem apresentar diferentes estágios de maturação. Um detector automático de bagas capaz de classificar o nível de maturação de frutos forneceria uma melhor avaliação da maturação da produção, determinante para tomada de decisão quanto ao melhor momento de colheita.

O presente trabalho propõe um método para detecção automática de bagas em imagens de cafeeiros tomadas em campo sob luz ambiente. De maneira similar aos trabalhos de Nuske et al. (2011) e Roscher et al. (2014) em viticultura, o presente método é composto por uma etapa de detecção de regiões candidatas na imagem, explorando o formato circular dos frutos de café observados, seguida por uma etapa de classificação supervisionada que visa eliminar os *falsos positivos* obtidos inicialmente.

MATERIAL E MÉTODOS

Imagens de cafeeiros foram tomadas na Fazenda Tozan (Monte d’Este), Campinas, em abril de 2015, utilizando-se uma câmera digital SLR (Canon® EOS Rebel T3i) com lentes 18–55 mm configurada em foco automático. Imagens coloridas de 8 bits contendo 5184×3456 pixels foram tomadas de plantas do cultivar Catuaí a uma distância de aproximadamente 1 m.

O método proposto é dividido em duas etapas. Na primeira, *regiões candidatas* são identificadas como possíveis localizações de frutos. Em seguida, um *classificador* é empregado para determinar se a região contém um fruto. Embora o classificador possa ser empregado a cada pixel da imagem, realizando assim uma varredura completa em todas as possíveis localizações, tal procedimento teria um alto custo computacional. Logo, é preferível reduzir o espaço de busca através de um procedimento capaz de filtrar o número de candidatos, descrito a seguir.

Detecção de candidatos

A detecção de candidatos visa encontrar regiões na imagem que são possíveis localizações de bagas de café. Para permitir a detecção adequada mesmo de frutos verdes, cuja cor pode ser similar a da folhagem do cafeeiro, optou-se por detectores de forma capazes de localizar regiões circulares na imagem a partir somente da intensidade dos pixels.

Quatro detectores foram estudados, entre eles a Transformada Circular de Hough (YUEN et al., 1990), também utilizada por Roscher et al. (2014) na detecção de bagas de uva. Os outros três detectores são conhecidos por identificar *blobs*, regiões mais claras cercadas por áreas mais escuras (ou vice-versa). São eles: Determinante da Hessiana (DoH), Laplaciano da Gaussiana (LoG) e Diferença de Gaussianas (DoG) (TUYTELAARS; MIKOLAJCZYK, 2007). Enquanto a Transformada Circular de Hough procura explicitamente por circunferências na imagem, os detectores de *blobs* identificam regiões salientes, isto é, que contrastam com o seu entorno, sendo suas dimensões limitadas pelo desvio padrão do núcleo de convolução utilizado.

Nessa etapa, o mais importante é limitar o número de falsos negativos, ou seja, bagas não identificadas. Já os falsos positivos, regiões erroneamente confundidas com bagas, são filtrados na etapa seguinte. Tradicionalmente, os erros de detecção são caracterizados pelas medidas de *precisão* e *cobertura* (*recall*). Sejam n_{fp} , n_{tp} e n_{fn} os números de falsos positivos, verdadeiros positivos e falsos negativos, respectivamente. Precisão e cobertura são definidas por:

$$\text{precisão} = \frac{n_{tp}}{n_{tp} + n_{fp}}, \quad (1)$$

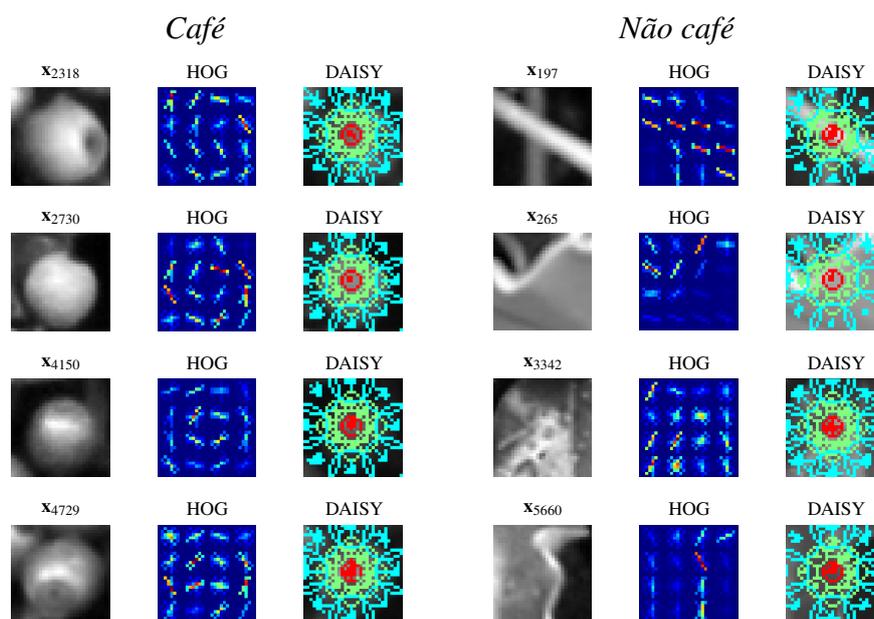
$$\text{cobertura} = \frac{n_{tp}}{n_{tp} + n_{fn}}. \quad (2)$$

Para determinar os valores de n_{fp} , n_{tp} e n_{fn} , foi realizada a anotação manual das bagas contidas em uma subimagem de 700×700 pixels, na qual foram identificadas 149 bagas visíveis. A partir dessa referência, os quatro detectores puderam ser avaliados com as medidas de precisão e cobertura.

Como será discutido na seção Resultados e Discussões, o detector LoG apresentou maior cobertura e foi utilizado para produzir um conjunto de 6.276 regiões candidatas. Todos os candidatos foram classificados manualmente, identificando-se 510 bagas de café e 5.766 falsos positivos. Essas regiões candidatas, manualmente classificadas e que chamaremos de *amostras*, foram empregadas no treinamento e avaliação de classificadores automáticos.

O conjunto de amostras utilizado na etapa de aprendizado consiste em 510 amostras da classe *café* e 2.883 amostras da classe *não café*, a metade dos falsos positivos, o que equilibra o número de amostras de cada classe. Esse conjunto foi dividido aleatoriamente em um conjunto de amostras para treinamento, X_{trein} , contendo 70% das amostras, e um conjunto de testes, X_{test} com 30% dos elementos. A Figura 1 exibe algumas amostras das duas classes. É importante notar que as amostras presentes na classe *não café* correspondem a partes da folhagem, galhos, buracos e especularidades, são essas as principais causas de falsos positivos encontrados por todos os detectores testados. O detector devolve, para cada candidato i , seu centro (x_i, y_i) na imagem e um raio r_i que determina o tamanho da região (idealmente, o raio da baba). Para cada amostra, uma janela quadrada, com $1.2 \cdot r_i$ de lado, é definida ao redor de (x_i, y_i) , de forma a caracterizar a região na periferia do candidato, como pode ser visto na Figura 1. Para obter uniformidade de escala, já que o valor de r_i é diferente para cada amostra, a janela passa por

Figura 1: Algumas amostras das classes *café* e *não café* e seus descritores HOG e DAISY.



uma transformação de escala, sendo redimensionada para 32×32 pixels por interpolação linear.

Descritores

Para cada amostra, descritores foram computados e utilizados na fase de classificação. Dois descritores foram empregados: HOG e DAISY. Os Histogramas de Gradientes Orientados (HOG) foram propostos por Dalal e Triggs (2005) e tornaram-se um descritor comumente utilizado em reconhecimento de objetos. A ideia é representar a aparência de um objeto pela distribuição da direção de suas bordas, na forma de um histograma da orientação dos gradientes na imagem. Esse descritor apresenta certa robustez a variações geométricas e fotogramétricas, de forma que pequenas rotações e translações têm pouco impacto no descritor se forem menores que a quantização utilizada no histograma. Para as amostras de café, cada imagem de 32×32 pixels foi dividida em uma grade de 16 células com 8×8 . Um histograma foi computado para cada célula, avaliando a distribuição da orientação das bordas em 9 direções diferentes. O resultado é um descritor composto por 144 dimensões, resultado da concatenação dos 16 histogramas. A Figura 1 exibe algumas amostras das duas classes e visualizações de seus descritores, nas quais é possível identificar as orientações dominantes em cada célula da amostra.

DAISY, proposto por Tola, Lepetit e Fua (2010), também caracteriza a aparência de um objeto através das as orientações de suas bordas, representadas através de *mapas de orientação convoluídos (convolved orientation maps)*. Uma região circular na imagem é representada por um vetor com n_o valores, em que n_o é o número de orientações avaliadas. O descritor para uma imagem com centro em um ponto (x_i, y_i) é montado através da amostragem de diversas regiões circulares ao redor de (x_i, y_i) sobrepostas, em diferentes escalas, que formam um padrão similar a uma “margarida” (Figura 1), o que origina o nome adotado pelos autores: DAISY. Para cada

uma de nossas amostras, os parâmetros utilizados produziram 17 regiões circulares, uma região central rodeada por outras 16 regiões dispostas em dois anéis concêntricos, como pode ser visto na Figura 1 e com mais detalhes em Tola, Lepetit e Fua (2010). Em cada região circular, valores são produzidos para 8 orientações diferentes, totalizando assim um descritor com 136 dimensões. Idealmente, a região circular central deveria caracterizar o centro da baga, as regiões no primeiro anel a circunferência do fruto e o anel mais externo as regiões de fundo que rodeiam a baga, como pode ser observado na Figura 1 em vermelho, verde e ciano respectivamente.

Classificação

As amostras são finalmente representadas como M vetores com N dimensões. M é igual a 2.375 e 1.018 para os conjuntos de treinamento e teste, respectivamente e N é 144 quando os descritores HOG são empregados e 136 para quando descritores DAISY são utilizados. Os vetores são também normalizados de modo que o conjunto de dados tenha média zero e variância unitária¹. Diversos classificadores foram testados e os dois que apresentaram melhor desempenho são apresentados neste trabalho: Máquinas de Vetores de Suporte (SVM) e *Gradient Boosting* (GB) (FRIEDMAN, 2001; HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Implementação

Para os detectores DoH, LoG e DoG, foram utilizadas as implementações disponíveis na biblioteca *scikit-image* (WALT et al., 2014), enquanto que a implementação da Transformada de Circular Hough empregada foi fornecida pela biblioteca OpenCV. Já a biblioteca *scikit-learn* (PEDREGOSA et al., 2011) forneceu as implementações para SVM e GB utilizadas.

RESULTADOS E DISCUSSÃO

A Figura 2 exibe os resultados da detecção obtidos por LoG e Hough, que apresentaram os melhores resultados nos testes, exibindo os valores de precisão e cobertura obtidos em relação ao conjunto de 149 bagas. A anotação manual também determinou a área de cada baga, de forma que é possível avaliar se os detectores são capazes de delimitar a área dos frutos. Se o detector não for capaz de isolar apropriadamente as bagas, os descritores utilizados na etapa seguinte não produzirão representações adequadas, resultando em erros de classificação.

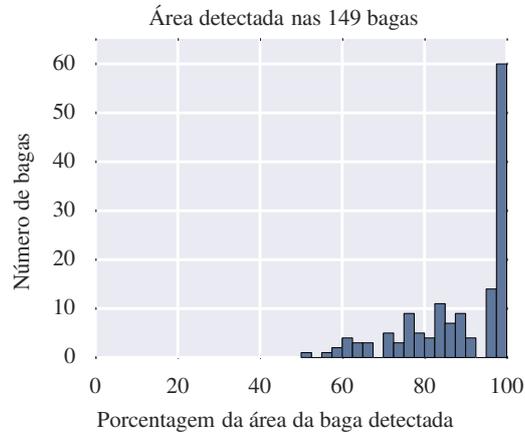
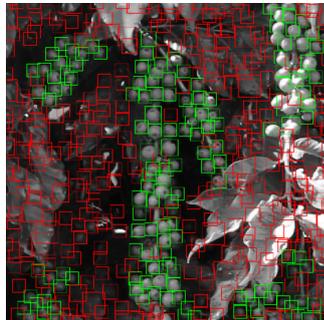
A medida de maior interesse na detecção de candidatos é a cobertura, pois é importante nessa etapa um baixo número de falsos negativos já que a fase de classificação é capaz de incrementar a precisão pela remoção de falsos positivos. LoG e a Transformada Circular de Hough apresentaram as melhores coberturas, 97% e 94% respectivamente². Porém, LoG oferece uma melhor cobertura da área dos frutos, resultando assim em melhores descritores para a fase de classificação (além de fornecer uma melhor estimativa do tamanho das bagas, caso essa medida seja de interesse). O que pode ser observado nos histogramas da Figura 2, que indicam que LoG conseguiu isolar mais de 90% da área do fruto para a maioria das bagas.

¹Este procedimento é comumente utilizado e evita que uma única dimensão, de maior escala relativa as outras, domine o processo de otimização.

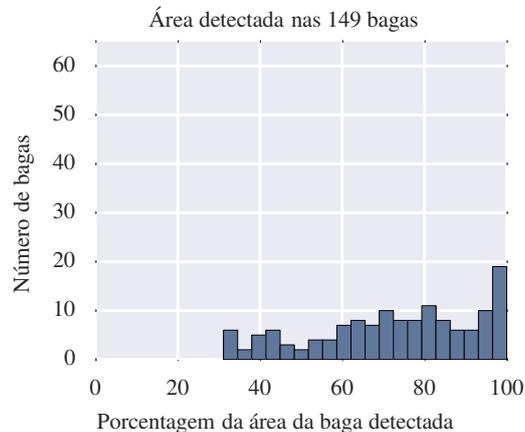
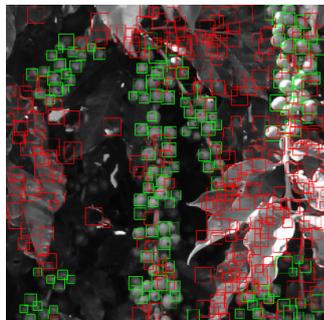
²DoH apresentou cobertura de 89% e DoG 82%.

Figura 2: Resultados da detecção de candidatos. Verdadeiros positivos e falsos positivos são exibidos em verde e vermelho respectivamente. Histogramas indicam a porcentagem da área da baga que foi encontrada pelo detector.

LoG - precisão: 0.32, cobertura 0.97



Hough - precisão: 0.32, cobertura 0.94



Os candidatos produzidos por LoG foram transformados em amostras e utilizados no treinamento e avaliação dos classificadores SVM e GB. Os resultados de precisão e cobertura para a classe *café* são exibidos na Tabela 1. Como observado por Nuske et al. (2011), a precisão final é o fator mais importante para a estimação de safra, assumindo que os níveis de cobertura e o número de bagas não visíveis permaneçam aproximadamente constantes ao longo do talhão. É possível ver na Tabela 1 que ambos os classificadores, SVM e GB, produzem resultados similares. Os descritores HOG produziram melhor precisão, mas a custo de uma perda de cobertura quando comparados aos descritores DAISY. Uma precisão de 90% foi atingida pelo uso de descritores HOG associados a um classificador SVM.

A Figura 3 exibe resultados de detecção de bagas para duas fotografias, *A* e *B*, obtidas em campo, utilizando-se a combinação HOG e SVM. Regiões classificadas como bagas, porém consideradas isoladas, são marcadas em amarelo, enquanto as demais são marcadas em azul. Os dados utilizados no treinamento e avaliação foram provenientes da fotografia *A*. O mesmo procedimento de detecção, incluindo os mesmos parâmetros e o mesmo classificador, sem novo treinamento, foi aplicado na imagem *B* e obteve bons resultados, um indício de sua robustez e

Tabela 1: Resultados para predição da classe *café* para o conjunto de teste.

Descritor	Classificador	Precisão	Cobertura
HOG	SVM	0,90	0,63
HOG	<i>Gradient Boosting</i>	0,89	0,65
DAISY	SVM	0,80	0,77
DAISY	<i>Gradient Boosting</i>	0,85	0,73

aplicabilidade.

CONCLUSÕES

O presente trabalho apresentou um método para detecção automática de frutos de café em imagens tomadas em campo com o uso de câmeras comuns. Este é possivelmente o primeiro estudo desse tipo para a cultura de café, e assemelha-se a estudos prévios encontrados na literatura para viticultura. Desse modo, é razoável esperar que o método proposto aqui possa também ser adaptado para viticultura e outras culturas nas quais os frutos apresentem padrões de bagas arredondadas, como uvas e mirtilos. A precisão da detecção obtida torna o método uma alternativa interessante para metodologias de predição de safra baseada em medidas fenológicas, embora um estudo mais amplo nessa direção ainda seja necessário.

Problemas com falsos positivos foram causados sobretudo por reflexão especular da luz. Uma metodologia mais apropriada para aquisição de imagens pode ser desenvolvida, determinando momentos do dia mais adequados à captura de imagens, o uso de anteparos para bloquear a luz solar direta ou outras alternativas. O cafeeiro é conhecido por ser uma planta que “esconde a produção”, já que a maioria dos frutos se encontra oculta na folhagem, principalmente em cultivares de café Arábica. Protocolos de aquisição de imagens também podem ser desenvolvidos, incluindo, por exemplo, procedimentos para exposição dos frutos pelo afastamento dos ramos.

Trabalhos futuros podem explorar novos descritores, a combinação de descritores e a redução de sua dimensionalidade. Outra área a ser explorada é a determinação de características fenológicas que habilitem formas mais sofisticadas de predição de safra por regressão, contrastando com metodologias atuais que se baseiam na composição de um único *índice fenológico* (OLIVEIRA, 2007; MIRANDA; REINATO; SILVA, 2014).

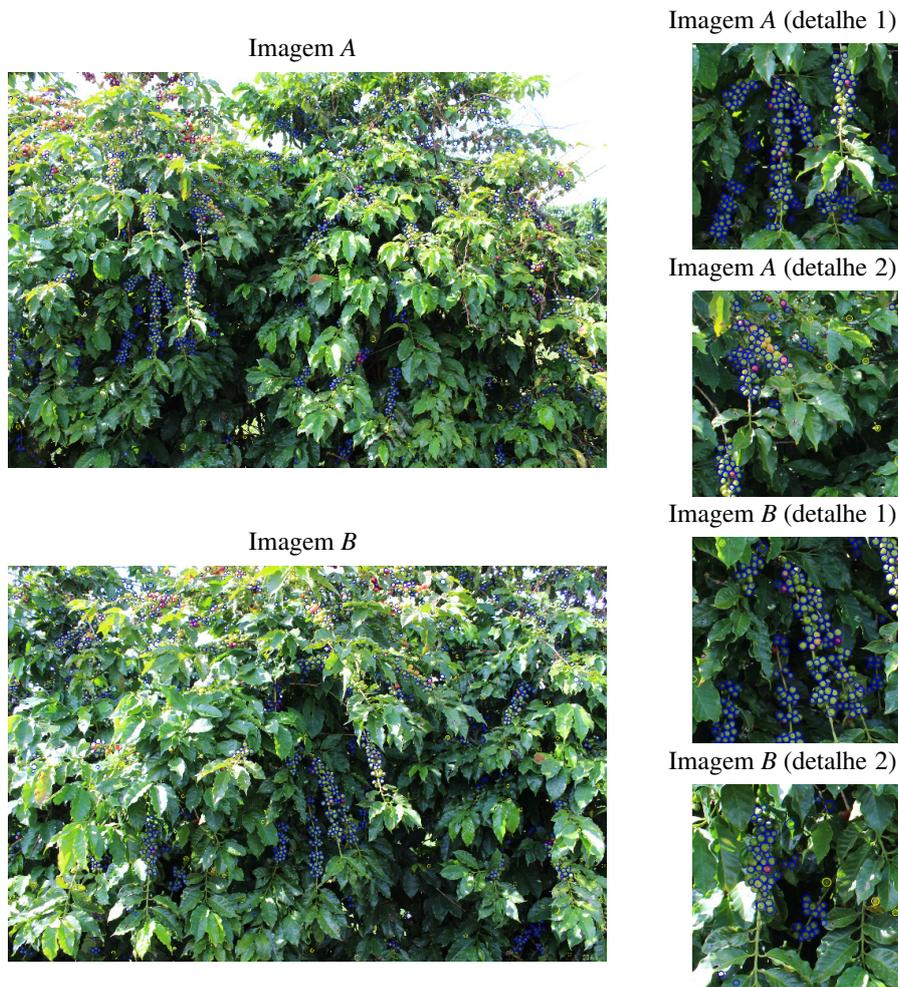
AGRADECIMENTOS

A pesquisa foi suportada pela Embrapa Informática Agropecuária. O autor agradece aos administradores da Fazenda Monte d’Este (Tozan) por permitirem a aquisição de imagens em seus cafezais. O autor também agradece a Flávio Regis de Arruda pelas discussões e sugestões sobre o emprego de descritores DAISY e o uso do algoritmo de *Gradient Boosting*.

REFERÊNCIAS

DALAL, N.; TRIGGS, B. Histograms of oriented gradients for human detection. *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, I, p. 886–893, 2005. ISSN 1063-6919.

Figura 3: Resultados para as imagens *A* e *B*.



FAHL, J. I. et al. Desenvolvimento e Aplicação de Metodologia para Estimativa da Produtividade do Cafeeiro. In: *Anais do Simpósio de Pesquisa dos Cafés do Brasil*. Brasília: Embrapa Café, 2005. p. CD-ROM. Disponível em: <<http://hdl.handle.net/10820/1412>>.

FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, JSTOR, p. 1189–1232, 2001.

FURBANK, R. T.; TESTER, M. Phenomics—technologies to relieve the phenotyping bottleneck. *Trends in plant science*, Elsevier Ltd, v. 16, n. 12, p. 635–44, dez. 2011. ISSN 1878-4372.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The elements of statistical learning: data mining, inference and prediction*. 2. ed. [S.l.]: Springer, 2009.

LAFFERTY, J. D.; MCCALLUM, A.; PEREIRA, F. C. N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: *Proceedings of the Eighteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan

- Kaufmann Publishers Inc., 2001. (ICML '01), p. 282–289. ISBN 1-55860-778-1.
- LEE, W. S. et al. Sensing technologies for precision specialty crop production. *Computers and Electronics in Agriculture*, v. 74, n. 1, p. 2–33, 2010. ISSN 01681699.
- LOY, G.; ZELINSKY, A. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 25, n. 8, p. 959–973, 2003. ISSN 01628828.
- MIRANDA, J. M.; REINATO, R. a. O.; SILVA, A. B. Modelo matemático para previsão da produtividade do cafeeiro. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 18, n. 4, p. 353–361, 2014.
- NUSKE, S. et al. Yield estimation in vineyards by visual grape detection. In: *IEEE International Conference on Intelligent Robots and Systems*. [S.l.: s.n.], 2011. p. 2352–2358. ISBN 9781612844541. ISSN 2153-0858.
- OLIVA, A.; TORRALBA, A. Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, v. 155 B, p. 23–36, 2006. ISSN 00796123.
- OLIVEIRA, A. d. O. *Estimativa da produção de café por meio de índice fenológico*. Dissertação (Mestrado) — Universidade Federal de Lavras, 2007.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- ROSCHER, R. et al. Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields. *Computers and Electronics in Agriculture*, Elsevier B.V., v. 100, p. 148–158, 2014. ISSN 01681699.
- SONKA, S. Big data and the ag sector: More than lots of numbers. *International Food and Agribusiness Management Review*, v. 17, n. 1, p. 1–20, 2014. ISSN 15592448.
- TOLA, E.; LEPETIT, V.; FUA, P. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Trans. Pattern Anal. Mach. Intell.*, IEEE Computer Society, Washington, DC, USA, v. 32, n. 5, p. 815–830, maio 2010. ISSN 0162-8828.
- TUYTELAARS, T.; MIKOLAJCZYK, K. Local Invariant Feature Detectors: A Survey. *Foundations and Trends® in Computer Graphics and Vision*, v. 3, n. 3, p. 177–280, 2007. ISSN 1572-2740.
- WALT, S. van der et al. Scikit-image: image processing in Python. *PeerJ*, v. 2, p. e453, 6 2014. ISSN 2167-8359.
- YUEN, H. et al. Comparative study of Hough Transform methods for circle finding. *Image and Vision Computing*, v. 8, n. 1, p. 71–77, 1990. ISSN 02628856.