



# 6 Visão computacional aplicada na agricultura

Thiago Teixeira Santos  
Jayme Garcia Arnal Barbedo  
Sônia Ternes  
João Camargo Neto  
Luciano Vieira Koenigkan  
Kleber Xavier Sampaio de Souza

## 1 Introdução

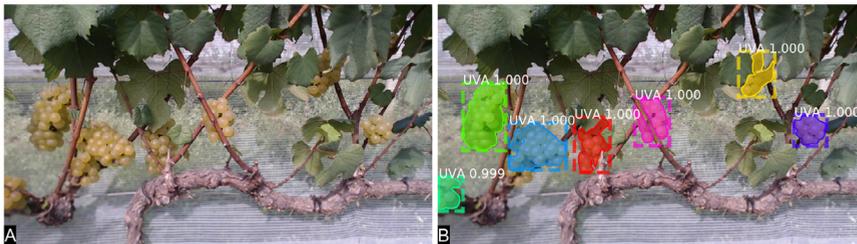
Em uma definição simples e abrangente, visão computacional é o campo da inteligência artificial dedicado à extração de informações a partir de imagens digitais. No contexto da agricultura digital, a visão computacional pode ser empregada na detecção de doenças e pragas, na estimação de safra e na avaliação não invasiva de atributos como qualidade, aparência e volume, além de ser componente essencial em sistemas robóticos agrícolas. Segundo Duckett et al. (2018), a robótica de campo deve viabilizar uma nova gama de equipamentos agrícolas: máquinas pequenas e inteligentes capazes de reduzir desperdício e impacto ambiental<sup>1</sup> e proporcionar viabilidade econômica, aumentando assim a sustentabilidade dos alimentos. Ainda segundo Duckett et al. (2018), há um potencial considerável no aumento da janela de oportunidades para intervenções, por exemplo, em operação em solos úmidos, operação noturna e monitoramento constante da lavoura.

Uma classe de problemas abordados pela visão computacional são os problemas ditos *perceptuais*: a detecção e a classificação de padrões, nas imagens, que são associados a um objeto de interesse, como frutos (Sa et al., 2016;

---

<sup>1</sup> Devido ao uso comedido e inteligente de defensivos ou simplesmente à intervenção mecânica: a remoção física de pragas por manipuladores.

Santos et al., 2020), animais (Barbedo et al., 2019) ou sintomas de doenças e pragas (Ferentinos, 2018; Barbedo, 2019). A partir de imagens capturadas por equipes de campo ou obtidas por câmeras acopladas em tratores, implementos, robôs ou drones, um monitoramento constante e eficiente pode ser realizado: a busca por anomalias na cultura ou na criação; a avaliação de variabilidade espacial da cultura para intervenção, segundo os preceitos da agricultura de precisão; e a atuação autônoma por máquinas e implementos. A Figura 1 exibe um exemplo: detecção de cachos de uva em imagens obtidas de vinhedos.



**Figura 1.** Exemplos de uma tarefa perceptual, a detecção de uvas em imagens: imagem tomada em um vinhola de uma videira Chardonnay (A); resultado da detecção com o uso de uma rede neural (B).

Ilustração: Thiago Teixeira Santos

Outra classe de problemas são os *geométricos*. Na formação de uma imagem, a luz captada pelas lentes é projetada em uma superfície, de modo que a cena tridimensional produz uma representação 2-D. Muito da estrutura da cena encontra-se na imagem, mas a informação de profundidade (a distância entre a câmera e os objetos em cena) é perdida. Uma das maiores contribuições da visão computacional geométrica foi o desenvolvimento de algoritmos capazes de recuperar a informação tridimensional perdida, a partir de um conjunto de imagens da mesma cena. Essa é uma das aplicações de visão computacional mais extensivamente utilizadas comercialmente na atualidade: o mapeamento tridimensional e a produção de mapas a partir de imagens obtidas por *Veículos Autônomos Não Tripulados* (VANTs – popularmente conhecidos como *drones*, ver a Figura 2). Metodologias baseadas em visão computacional geométrica têm sido empregadas em estudos geológicos (Westoby et al., 2012), na avaliação de altura de pastagens (Forsmo et al., 2018) e no mapeamento de culturas (Comba et al., 2018), entre outros usos. Comercialmente, é a tecnologia principal por trás de serviços de mapeamento e reconstrução 3-D por VANTs extensivamente utilizados na agricultura, como o *Pix4D mapper* e o *Agisoft PhotoScan/Metashape*.

Na pesquisa agropecuária, há um número crescente de aplicações de visão computacional. Considere como exemplo o periódico *Computers and Electronics in Agriculture*, especializado em novas aplicações de software, hardware e eletrônica na agricultura. Uma busca por artigos relacionados à visão computacional revela que 23,7% de todos os trabalhos publicados em 2018 são associados ao tema, subindo para 29,1% em 2019. De janeiro

**Figura 2.**

Mapeamento com VANT: as imagens são utilizadas para identificar a estrutura tridimensional da área e a posição e a orientação da aeronave, exibida em vermelho (A); o modelo tridimensional geolocalizado é então projetado em um plano, formando um mapa (B).

Ilustração: Thiago Teixeira Santos



a junho de 2020, 115 dos 319 trabalhos (36,0%) publicados são associados à visão computacional. Tal volume de artigos também se traduz em impacto: dos 25 trabalhos mais citados até junho de 2020, 14 são aplicações de visão computacional. Alguns fatores simples explicam esse crescimento. Câmeras digitais são dispositivos acessíveis e largamente disponíveis em diversas configurações, facilmente integráveis a sistemas maiores (como *smartphones* e

VANTs). Avanços em algoritmos e hardware surgidos nos últimos dez anos viabilizaram o atual vigor da área.

Nas próximas seções apresentaremos avanços recentes na aplicação de visão computacional à agricultura, com foco em contribuições realizadas nos últimos três anos pela Embrapa Informática Agropecuária. Esses avanços são oriundos tanto da *visão computacional perceptual*, o reconhecimento dos elementos na cena (Seção 2), quanto da *visão computacional geométrica*, a recuperação de informação tridimensional a partir de imagens (Seção 3). A combinação de ambas as frentes (Seção 4) abre caminho para sistemas capazes de realizar operações de alta complexidade, como a robótica de campo. A Seção 5 fecha o capítulo com algumas considerações finais.

## 2 Percepção: reconhecimento de padrões em imagens

O reconhecimento de padrões pode ser visto como a tarefa de encontrar uma *representação* para o padrão procurado que seja versátil, de modo a cobrir as variações observáveis, mas simples o suficiente para ser processada em tempo hábil pela máquina. Em outras palavras, trata-se de uma descrição do padrão suficiente para permitir que a máquina o encontre nos dados de entrada, porém sucinta de modo que sua interpretação seja realizada dentro de restrições de tempo de operação.

Padrões visuais em imagens naturais podem ser incrivelmente intrincados, com regularidades e variações difíceis de descrever. Na agricultura, padrões assumidos por frutos, folhas, grãos, plantas e sintomas de patologias apresentam uma enorme variabilidade, ampliada pelas diferenças de iluminação, posição, oclusão e fontes diversas de ruído (lentes sujas, poeira, interferência, etc.). A Figura 3 ilustra algumas dessas dificuldades enfrentadas por um sistema de detecção de frutos em condições reais de cultivo em campo: oclusão severa entre frutos, folhas e galhos; similaridade de cor entre frutos verdes e o dossel; variações de iluminação entre imagens; reflexão especular (reflexo direto da luz do sol que satura o sensor da câmera); e problemas de foco. Apesar de algum sucesso decorrente do emprego de técnicas de aprendizado de máquina (Gongal et al., 2015), o reconhecimento de padrões em imagens naturais começou a atingir níveis elevados de acurácia com o advento das *redes neurais convolutivas* (Lecun et al., 2015), rapidamente adotadas no reconhecimento de imagens na agricultura (Kamilaris; Prenafeta-Boldú, 2018).

Em redes neurais, uma *arquitetura* ou *modelo* é uma sequência de módulos que realizam operações simples nos dados, de maneira que um módulo recebe dados dos módulos anteriores e *propaga* o resultado de suas operações para os módulos seguintes. Em visão computacional, as redes neurais mais utilizadas

**Figura 3.**

Exemplos das dificuldades enfrentadas na detecção de frutos. Nas imagens, podemos observar problemas de foco, reflexão especular, oclusão severa por folhas, galhos e outros frutos, variações de luminosidade e similaridades no padrão de cores entre frutos e folhas.



Fonte: Adaptado de Camargo Neto et al. (2019) a partir de material cedido pela PES/Fundecitrus.

são as redes neurais convolutivas (*convolutional neural networks* – CNNs), nas quais a principal operação empregada é a *convolução*, uma combinação linear dos valores na vizinhança dos pixels de entrada. Redes neurais são ditas *profundas* se há uma sequência grande de módulos encadeados. Quanto mais profunda a rede, maior sua capacidade de aprender representações para padrões complexos, uma vez que cada módulo é capaz de compor as representações dos módulos anteriores em uma hierarquia. No caso de imagens, há uma interpretação intuitiva para esse comportamento: os módulos iniciais são capazes de encontrar linhas e bordas dos objetos, os módulos seguintes compõem esses padrões em texturas e estruturas simples como triângulos e manchas, que são então combinadas em outras estruturas como partes de folhas, galhos e bagas. Finalmente, os módulos finais combinam esses elementos em objetos de interesse: uma planta, um cacho de uvas, um boi.

Os módulos possuem *parâmetros* que precisam ser ajustados de modo que a operação conjunta de toda a rede produza os resultados esperados. Uma metáfora frequentemente utilizada é imaginar que cada parâmetro é ajustado por um *dimmer*. Ajustar uma rede neural seria realizar o ajuste de *milhões* de *dimmers*, cada um deles podendo afetar o desempenho do reconhecimento dos padrões. Manualmente, tal ajuste seria impraticável e virtualmente impossível. O *treinamento* de redes neurais é justamente um processo automatizado para o ajuste desses parâmetros, de modo que a rede “aprenda” as representações apropriadas ao problema de reconhecimento em questão.

No aprendizado supervisionado de padrões em imagens, esse treinamento é realizado com o uso de *observações*, imagens cuja resposta desejada é conhecida (“há uma laranja nesta imagem”, “há sinais de ferrugem do café nesta folha”). Milhares de observações são necessárias para o treinamento, algo diretamente ligado ao tamanho da rede: mais parâmetros exigem mais observações, embora seja difícil estabelecer uma relação exata entre o número

de parâmetros e o número de observações necessárias. Quando a rede processa a imagem de entrada, o resultado produzido é comparado ao resultado esperado, e o erro entre eles é computado. Os parâmetros são então ajustados de modo a diminuir o erro anterior, em um processo conhecido como *backpropagation* (Goodfellow et al., 2016). Na prática, as observações são agrupadas em *lotes* (*batches*), a rede processa o lote e o erro observado é computado. O algoritmo de *backpropagation* é utilizado para ajustar os parâmetros, iniciando pelos módulos finais da rede e prosseguindo em direção aos parâmetros dos módulos iniciais (daí se origina o nome do procedimento). O treinamento prossegue com o lote seguinte, e o procedimento é repetido até que o erro atinja um mínimo observável<sup>2</sup>. Em suma, redes neurais profundas automatizam o processo de busca por representações adequadas em problemas de reconhecimento de padrões, desde que haja um conjunto suficientemente grande de observações para treinamento que representem adequadamente a variabilidade do padrão pretendido. É justamente essa capacidade que torna a metodologia tão atrativa aos problemas intrincados de reconhecimento encontrados na agricultura.

## 2.1 Identificação de doenças em plantas

A detecção e a classificação de sintomas de doenças, pragas e deficiências nutricionais de plantas em imagens são de enorme interesse na agricultura. A detecção automática possibilita o monitoramento constante e a busca por anomalias na cultura, a partir de imagens capturadas por equipes de campo ou obtidas por câmeras acopladas em tratores, implementos, robôs ou VANTs. Já a classificação associa as anomalias detectadas à doença, deficiência ou praga, auxiliando o produtor na intervenção correta. Redes neurais podem ser empregadas em ambas as tarefas, inclusive simultaneamente.

Como visto anteriormente, milhares de observações são necessárias até que uma rede neural seja capaz de produzir resultados acurados. Essa necessidade é ampliada no reconhecimento de doenças em plantas devido ao enorme número de combinações resultante do cruzamento entre culturas alvo, patologias, estágio de desenvolvimento da doença e condição de imageamento (coleta manual, monitoramento aéreo por VANTs, captura no nível do solo por máquina, pose da câmera, entre outras). Esse cenário aponta para a necessidade de grandes bases de dados compartilhadas (Barbedo, 2018; Ferentinos, 2018), uma vez que sua produção exige considerável esforço.

---

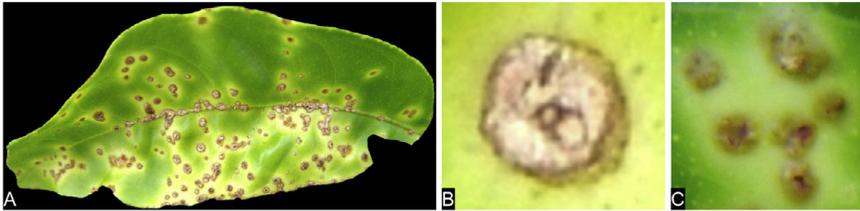
<sup>2</sup> O erro ideal seria zero, mas não há garantia de que uma arquitetura seja capaz de atingi-lo. Também é um problema em aberto determinar *a priori* qual o menor erro que uma rede será capaz de atingir para um certo conjunto de treinamento.

O processo de coleta e anotação das imagens, isto é, a associação de cada imagem com o resultado desejado para a etapa de aprendizado supervisionado, geralmente é laborioso e custoso. Porém, algumas estratégias podem ser empregadas para aumentar o número de observações. Barbedo (2019) mostrou que o fato de múltiplas lesões das mesmas patologias ocorrerem em uma mesma folha pode ser explorado para incrementar o número de observações oriundas de uma mesma coleta. Diversos exemplos de sintomas podem ser obtidos de uma única folha ou amostra de tecido vegetal, como pode ser observado na Figura 4. Essa estratégia permitiu que uma base de dados original, contendo 1575 observações (Barbedo, 2018), fosse expandida para 46409 observações (Barbedo, 2019), produzindo ganhos de acurácia na classificação de doenças de, em média, 12%.

**Figura 4.**

Exemplos de observações utilizadas no treinamento de sistemas para reconhecimento de doenças em plantas: uma amostra de uma folha doente, coletada em campo (A); uma observação de sintoma associado à patologia (B); aglomerados de sintomas que também formam um padrão discernível associável à patologia (C).

Fonte: Adaptado de Barbedo (2019).



Barbedo (2019) mostrou que uma rede neural convolutiva, a arquitetura GoogLeNet (Szegedy et al., 2015), pode ser aplicada na classificação de um grande número de patologias em diversas culturas, atingindo valores de acurácia de 80% (maracujá) a 100% (mandioca, couve, algodão, trigo e cana-de-açúcar), como pode ser observado na Tabela 1. A base de dados utilizada, batizada como *Digipathos*, foi disponibilizada publicamente<sup>3</sup>. Embora os resultados da classificação sejam promissores, grandes desafios ainda persistem, principalmente em respeito à detecção (“há sintomas presentes na observação?”), crucial no monitoramento autônomo para manejo de doenças e pragas, mas que ainda não apresenta a mesma acurácia da classificação (“qual a patologia para o sintoma observado?”). Em seus experimentos, Barbedo (2019) mostra que detecções acuradas podem ser produzidas quando os sintomas já são severos, mas o mesmo não ocorre quando os sintomas ainda são brandos ou não ocupam grandes porções do tecido vegetal, justamente o momento ideal para intervenção pelo agricultor. Erros de detecção do tipo *falso positivo* (tecido saudável detectado como doente) são frequentemente causados por fatores como presença de poeira, detritos ou mesmo gotículas de água. Também não é claro ainda qual o número de amostras necessário para que as características dos sintomas sejam propriamente aprendidas

<sup>3</sup> Disponível em: <https://www.digipathos-rep.cnpia.embrapa.br>.

Cultura	Número de imagens	Acurácia (%)
Feijão	3.079	94 ± 0,8
Mandioca	895	100 ± 0,0
Citros	1.868	96 ± 0,6
Coco	1.504	98 ± 0,6
Milho	10.480	75 ± 4,4
Café	1.899	89 ± 1,9
Algodão	2.023	99 ± 0,3
Caju	4.509	98 ± 0,5
Uva	2.330	96 ± 0,8
Couve	196	100 ± 0,0
Maracujá	280	80 ± 4,2
Soja	13.733	87 ± 3,6
Cana-de-açúcar	2773	99 ± 0,4
Trigo	840	99 ± 0,5
<i>Total</i>	46.135	94 ± 2,0

**Tabela 1.** Acurácia de classificação de patologias em diversas culturas. Para as imagens de mandioca e couve, a acurácia alcançou 100% em todos os testes.

Fonte: Adaptado de Barbedo (2019).

pelos redes neurais (uma questão ainda em aberto na visão computacional de forma geral).

## 2.2 Detecção de animais em pastagens

Barbedo et al. (2019) apresentam um exemplo de como as tecnologias de VANTs e a visão computacional podem ser combinadas para possibilitar o monitoramento de enormes áreas, por exemplo na detecção de gado na pecuária extensiva. Os pecuaristas têm grandes dificuldades de acompanhar os rebanhos nas pastagens, devido à dinâmica dos animais e à enorme dimensão das áreas de pastagem.

Uma base de dados composta por 1853 imagens contendo 8629 animais da raça Canchim foi produzida a partir de imagens obtidas por um quadricóptero disponível comercialmente<sup>4</sup>. Barbedo et al. (2019) testaram 15 arquiteturas diferentes de redes neurais em 3 resoluções espaciais distintas (1, 2 e 4 cm/pixel),

<sup>4</sup> No caso, um veículo DJI Phantom 4 Pro.

de forma a analisar o desempenho decorrente de diferentes alturas de voo. Os resultados mostraram que a maioria das arquiteturas testadas conseguiram atingir níveis elevados de acurácia, superiores a 95%. A arquitetura NasNet (Zoph et al., 2018), uma rede muito profunda e de grande capacidade de aprendizado de padrões complexos, atingiu acurácias próximas a 100%. Esses resultados são expressivos, especialmente se considerada a complexidade do problema, como pode ser visto na Figura 5: diversas situações, de oclusão severa por árvores e bebedouros a diferenças de iluminação e condição da pastagem, além da própria posição e disposição dos animais, apresentam uma enorme variabilidade de situações. Mesmo assim, a acurácia da maioria das arquiteturas testadas é expressiva. Outro efeito particularmente interessante do ponto de vista operacional foi o fato de a maioria dos modelos apresentar melhores resultados na resolução de 2 cm/pixel e não na resolução máxima de 1 cm/pixel, algo possivelmente decorrente da resolução dos módulos convolutivos com que essas arquiteturas foram projetadas originalmente. Na prática, isso viabiliza voos em alturas mais elevadas, o que permite a cobertura das aéreas em menos tempo.

**Figura 5.** Exemplos de situações observadas na detecção de animais em pastagens: animal em pasto alto (A); pasto seco (B); solo exposto (C); oclusões por árvores (D); coberturas de bebedouros (E) e cabos elétricos (F).

Fonte: Adaptado de Barbedo et al. (2019).



### 2.3 Detecção e contagem de frutos

A detecção automática de frutos é um componente habilitador para diversas aplicações em agricultura. Ela pode auxiliar a estimativa de produção, útil no planejamento logístico e na negociação entre o produtor rural e os compradores. Se a detecção é combinada à localização precisa no espaço, novas aplicações em agricultura de precisão podem ser desenvolvidas, auxiliando

no gerenciamento adequado da variabilidade espacial nas lavouras. A detecção de frutos também pode ser uma etapa preliminar para o monitoramento de doenças e deficiências nutricionais (vistas na Seção 2.1), restringindo as áreas nas imagens que devem ser inspecionadas em busca de sintomas. Considerando o declínio na força de trabalho na agricultura, a detecção de frutos também é uma tecnologia habilitadora para sistemas de pulverização e colheita automatizados (Duckett et al., 2018; Xiong et al., 2020).

Como discutido anteriormente, diversos fatores dificultam o processo de detecção, de oclusão por folhas e galhos a problemas de foco da câmera e iluminação (Figura 3). Em algumas culturas, os frutos apresentam ainda uma diversidade de formatos, compacticidade e orientação, como no caso da viticultura (Santos et al., 2020). Apesar de algum sucesso obtido com outras técnicas de aprendizado de máquina (Gongal et al., 2015), a detecção de frutos realmente ganhou força recentemente, com os avanços em redes neurais convolutivas (Sa et al., 2016; Bargoti; Underwood, 2017; Kamilaris; Prenafeta-Boldú, 2018).

Camargo Neto et al. (2019) produziram uma base de dados com 3066 imagens de laranjas coletadas em campo, oriundas de diferentes dispositivos, como câmeras fotográficas e *smartphones*. A maior parte das imagens foi fornecida pelo Programa de Estimativa de Safra (PES) do Fundo de Defesa da Citricultura (Fundecitrus). Os frutos, de diversas variedades de laranja, apresentavam diferentes níveis de maturação, com predominância de frutos verdes (Figura 3). Dessas imagens, um subconjunto de 2036 observações foi utilizado no treinamento de uma rede neural YOLOv3 (Redmon et al., 2016; Redmon; Farhadi, 2018). Ao avaliarem a rede treinada nas 1030 imagens restantes, os autores apuraram a correta detecção de mais de 90% dos frutos, com uma precisão também acima de 90%, ou seja, menos de 10% das detecções produzidas eram falsos positivos. A Figura 6 exibe um exemplo da detecção de frutos em uma imagem de laranjeira obtida em campo.

Santos et al. (2020) mostraram que, para uvas em viticultura, culturas que apresentam grande variabilidade em forma, cor, tamanho e compactidade, os cachos podem ser detectados e segmentados com o uso de arquiteturas como Mask-RNN e YOLO. Os autores produziram uma nova ferramenta de anotação capaz agilizar o processo de associação de pixels a frutos, discriminando exatamente quais pixels pertencem a quais cachos. A base de dados gerada, batizada de WGISD (*Embrapa Wine Grape Segmentation Dataset*) e disponibilizada publicamente<sup>5</sup>, contém 4432 cachos em 300 imagens, cobrindo cinco variedades vinícolas. Os autores avaliaram três diferentes arquiteturas de redes neurais, YOLOv2 (Redmon; Farhadi, 2017), YOLOv3 (Redmon;

<sup>5</sup> Disponível em: <https://doi.org/10.5281/zenodo.3361736>.

**Figura 6.**

Resultados de detecção de laranjas em campo com o uso de uma rede neural convolutiva YOLOv3: uma laranjeira imageada em campo, os frutos detectados pela rede em vermelho (A); detalhe (B).

Fonte: Adaptado de Camargo Neto et al. (2019) a partir de material cedido pela PES/Fundecitrus.



Farhadi, 2018) e Mask-RCNN (He et al., 2017), esta última responsável pelos resultados mais promissores. Em uma base de teste composta por 837 cachos, a rede identificou 87% dos cachos, com uma precisão de 90.7%. Exemplos das detecções produzidas são exibidos na Figura 1 (B).

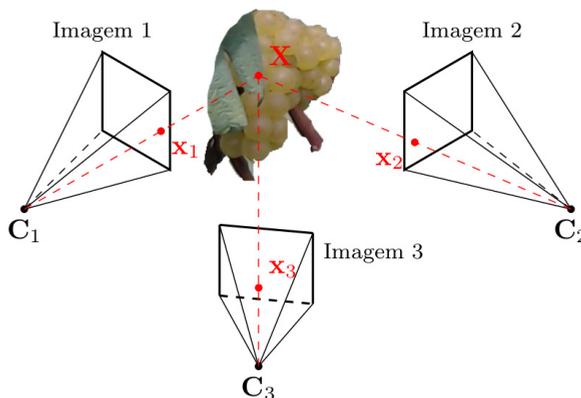
Contudo, uma aplicação completa de contagem de frutos necessita de uma metodologia capaz de integrar as detecções reportadas em várias imagens, de forma que frutos vistos em mais de uma imagem não sejam contabilizados múltiplas vezes. Em outras palavras, frutos (e objetos de interesse em geral) observados em diversas imagens devem ser *associados* entre si. Essa tarefa de *associação de dados* pode ser desempenhada pela integração do reconhecimento de padrões com a visão computacional geométrica, como veremos a seguir.

### 3 Mapeamento e reconstrução tridimensional

Uma das maiores contribuições da visão computacional geométrica foi o desenvolvimento de algoritmos capazes de recuperar a informação tridimensional a partir de um conjunto de imagens da mesma cena. Resultados de décadas de pesquisa em áreas como geometria projetiva e otimização contínua, esses algoritmos são capazes de transformar até uma simples *webcam* em um

poderoso *scanner* 3-D. Talvez ainda mais importante, eles permitem a um agente móvel, como um VANT, não só mapear a estrutura tridimensional do ambiente, mas também determinar sua localização precisa no local (Figura 2), abrindo caminho para agentes autônomos capazes de navegar e interagir com seu entorno (Stachniss et al., 2016).

As imagens devem ser obtidas de posições diferentes, por múltiplas câmeras ou por uma única câmera que se *movimenta* pela cena. Esse é o significado do termo *structure from motion* – *SfM* (estrutura a partir do movimento), utilizado em visão computacional para definir o problema de recuperação da estrutura tridimensional de uma cena e da posição da câmera a partir de um conjunto de imagens. A Figura 7 ilustra o processo de projeção de um ponto na cena, conforme a câmera é movida para três posições diferentes. Se pudermos determinar correspondências entre pontos nas diferentes imagens, é possível estabelecer, com o auxílio de técnicas de geometria projetiva, a posição da câmera no momento da captura de cada imagem, mais precisamente a localização de seus *centros de projeção*, representados na Figura 7 pelos pontos  $C_1$ ,  $C_2$  e  $C_3$ <sup>6</sup>. Uma vez estabelecida a localização dos centros de projeção, é possível então estimar a posição do ponto no espaço tridimensional a partir de suas projeções nas imagens (os pontos  $x_1$ ,  $x_2$  e  $x_3$  na Figura 7), um processo conhecido como *triangularização*. Uma descrição detalhada de todo o processo pode ser vista em Hartley e Zisserman (2003). A determinação de correspondências em imagens também é obtida automaticamente, com o uso de algoritmos especializados em encontrar pontos visualmente salientes (os pontos) e, pela comparação dos pixels em suas vizinhanças, associar pontos de imagens diferentes (Lowe, 2004; Detone et al., 2018).



**Figura 7.** *Structure from motion*. Um ponto  $X$  em uma superfície da cena é projetado no plano de imagem em posições diferentes conforme a câmera é movida para as posições  $C_1$ ,  $C_2$  e  $C_3$ . Ilustração: Thiago Teixeira Santos.

<sup>6</sup> Informação adicional, obtida por métodos de calibração, é necessária para estabelecer a escala correta, isto é, a distância em uma unidade conhecida como metros ou milímetros.

Santos et al. (2017) mostraram que um sistema de SfM utilizando uma simples webcam pode construir modelos tridimensionais acurados de plantas em campo. A Figura 8 exhibe um exemplo, para uma videira de *Chardonnay*. Como veremos na próxima seção, esses modelos tridimensionais podem ser empregados na estimação de atributos 3-D, como volume de frutos e posição. O sistema 3-D utilizado e desenvolvido na Embrapa Informática Agropecuária, batizado de *3dmcap*, encontra-se disponível gratuitamente<sup>7</sup> para uso não comercial.

**Figura 8.**

Reconstrução tridimensional com SfM.

Ilustração: Thiago Teixeira Santos



O uso de informação tridimensional na agricultura deve se intensificar nos próximos anos, não só pelo uso da técnica de SfM (já comercialmente empregada por serviços de mapeamento 3-D com VANTs), como pelo barrateamento de câmeras estéreo, que fornecem informação de profundidade na imagem, e de sensores LIDAR. Exemplos recentes são o uso de câmeras estéreo na fenotipagem de vinhedos (Milella et al., 2019) e a detecção de maçãs com o uso de LIDAR (Gené-Mola et al., 2020).

Na Figura 8 temos a reconstrução tridimensional com SfM para uma videira de Chardonnay em campo: os prismas vermelhos indicam a posição e a orientação da câmera (uma webcam comercial, no momento da captura de cada imagem (A); mesmo modelo 3-D observado de outro ângulo (B).

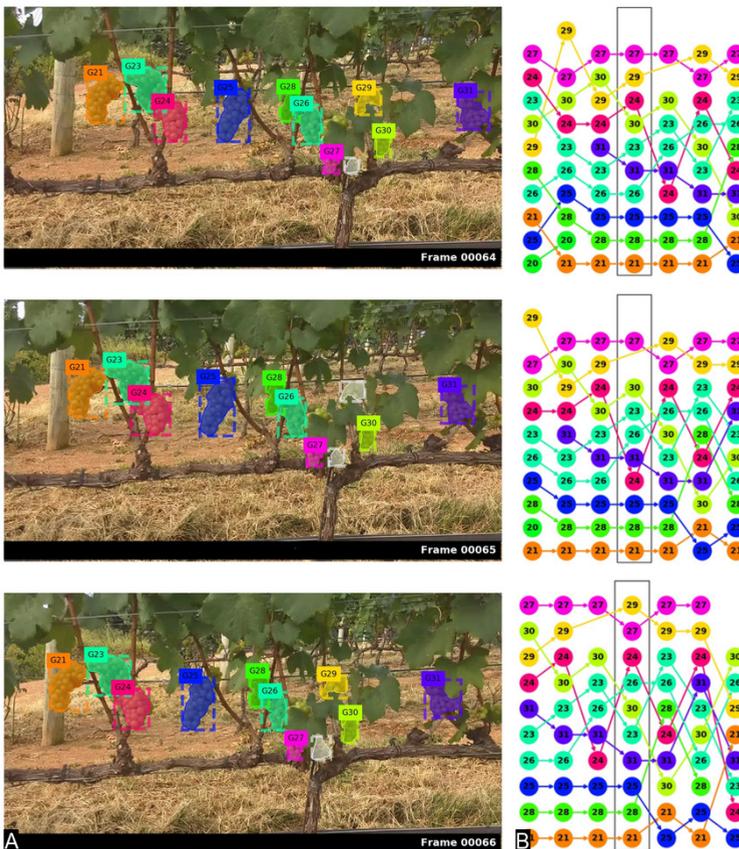
## 4 Combinação de estrutura e reconhecimento

Se o SfM recupera a estrutura tridimensional da cena e do próprio imageamento (a posição da(s) câmera(s) no momento da captura), e o reconhecimento identifica objetos de interesse na cena, como sintomas, frutos, plantas ou animais, a combinação das duas informações possibilita uma avaliação ampla do ambiente observado.

Um dos usos dessa combinação é o mapeamento de frutos: a informação 3-D combinada com a detecção de frutos em cada imagem permite que a posição espacial de cada fruto seja determinada e que o mesmo fruto não seja

<sup>7</sup> Disponível em: <https://github.com/thsant/3dmcap>.

contabilizado mais de uma vez quando aparece em múltiplas imagens. Santos et al. (2020) empregaram SfM para obter uma reconstrução tridimensional de uma linha de videiras em campo, a partir dos quadros de uma sequência de vídeo produzida por uma câmera embarcada em um veículo de serviço. Como descrito na Seção 2.3, uma rede neural foi utilizada para detectar cachos de uvas em cada imagem. Pela projeção dos pontos 3-D da cena nas imagens, foi possível associar detecções a posições no espaço tridimensional e, com isso, estabelecer correspondências entre os cachos observados em um quadro do vídeo e os cachos vistos nos quadros seguintes<sup>8</sup> (Figura 9). A utilização conjunta de modelos 3-D obtidos por SfM e redes convolutivas para detecção e contagem de frutos também foi explorada por Liu et al. (2019) em pomares de manga e por Häni et al. (2020) em pomares de maçã.



**Figura 9.** Rastreamento de cachos de uva em uma sequência de vídeo obtida em campo: quadros do vídeo foram extraídos e submetidos à detecção de frutos por redes neurais (A); os nós representam cachos de uva, na ordem em que foram encontrados pela rede neural (cada coluna de nós representa um quadro da sequência de vídeo). As setas informam a associação entre nós de um quadro para outro, realizada com o uso de informação 3-D obtida por SfM (B).  
Fonte: Adaptado de Santos et al. (2020).

<sup>8</sup> Um vídeo demonstrando o rastreamento de cachos encontra-se disponível em: <https://www.youtube.com/watch?v=1Hji3GS4mm4>.

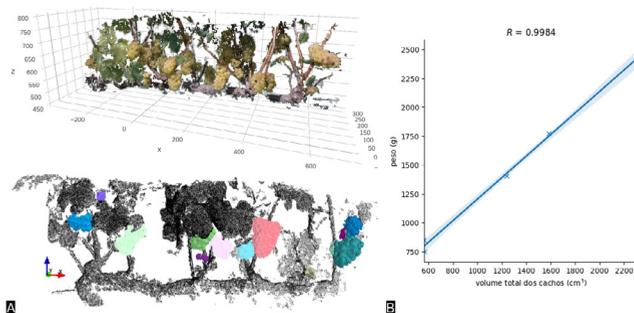
Atributos de grande interesse em aplicações agrônômicas podem ser extraídos da informação tridimensional. Santos et al. (2017) utilizaram um algoritmo de aprendizado de máquina para identificar quais regiões dos modelos tridimensionais de videiras correspondiam aos cachos de uva, como visto na Figura 10 (A). O volume dos cachos foi então estimado a partir dessas regiões. Volume do fruto apresenta forte correlação com seu peso, como pode ser observado na Figura 10 (B). Tais sistemas baseados em visão computacional podem fornecer uma metodologia não invasiva e não destrutiva de estimação do peso dos frutos, sem a necessidade de removê-los da planta. Tal tecnologia pode ser empregada na avaliação do crescimento ao longo do ciclo da cultura, sem a necessidade de remoção (colheita) de amostras.

Na Figura 10, vemos a estimação de peso de frutos a partir do volume em modelos tridimensionais: cachos de uva são identificados (em cores) e separados do restante da planta (em preto) (A); coeficiente de determinação entre o volume estimado e o peso total de frutos em cinco videiras diferentes (B).

**Figura 10.**

Estimação de peso de frutos a partir do volume em modelos tridimensionais.

Fonte: Adaptado de Santos et al. (2017).



#### 4.1 Atuação e intervenção: robótica de campo

A combinação de SfM e reconhecimento é justamente uma das tecnologias habilitadoras para uma das aplicações mais desafiadoras e impactantes na automação agrícola: a *robótica de campo*. Considere, como exemplo, um grande desafio da robótica agrícola: a *colheita automatizada* de frutas. Enquanto culturas como grãos, cana-de-açúcar e café contam com maquinário próprio para colheita automatizada, o mesmo não vale para a horticultura e fruticultura – especialmente para esta última –, devido à complexidade encontrada na estrutura dos pomares. A colheita de frutas depende de colheita manual, o que é preocupante se considerada a disponibilidade decrescente de mão de obra no campo (Roser, 2013).

Sistemas de colheita automática necessitam dos dois componentes da visão computacional: o perceptual, para identificação de frutos e obstáculos, e o geométrico, para posicionamento automático do robô e seus manipuladores. Diversos grupos de pesquisa têm aplicado esses dois componentes no

desenvolvimento de sistemas de colheita automatizada. Tomando a pomicultura como exemplo, foi desenvolvido um sistema robótico para colheita de maçãs, avaliado em um pomar comercial. Seu sistema de visão computacional era preciso, levando em média 1,5 s para localizar cada fruto. O sistema teve sucesso na colheita de 85% dos frutos, com um tempo médio de 6 s por fruto. Além da pomicultura, outras culturas têm sido investigadas para a implementação de colheita robotizada, como pimentões (Bac et al., 2017), alface (Birrell et al., 2020), morango (Xiong et al., 2020), kiwi (Williams et al., 2020), entre outras.

## 5 Considerações finais

A visão computacional tem um enorme potencial de aplicação no âmbito da agricultura digital. Diversos produtos e serviços baseados em componentes de visão computacional devem chegar aos produtores nos próximos anos. Contudo, muitos desafios ainda dependem de esforços de pesquisa e desenvolvimento.

Um grande gargalo é a necessidade de enormes bases de dados para o treinamento de redes neurais para tarefas perceptuais. Pesquisas na área de aprendizado semisupervisionado e não supervisionado vêm sendo conduzidas pela comunidade de visão computacional. A ideia é poder aprender padrões de interesse com poucos exemplos e obter sistemas de boa acurácia na detecção de padrões como frutos, sintomas e animais.

Na robótica, o desafio segue sendo o desenvolvimento de sistemas robustos, capazes de operar em campo por longos períodos com autonomia, mas que sejam seguros para pessoas e animais em circulação. Esses sistemas precisam mapear o ambiente rapidamente, responder prontamente, encontrar de forma acurada os objetos a serem monitorados e executar as intervenções para as quais foram designados. Apesar dos imensos desafios, as comunidades de visão computacional e robótica obtiveram grandes avanços nos últimos anos, que devem se refletir em breve em diversas aplicações agrícolas, do monitoramento à atuação.

Por fim, os autores ressaltam que os resultados em detecção de frutos foram financiados pelos projetos Embrapa SEG 11.14.09.001.05.04 e Fapesp 2017/19282-7. Os resultados relacionados à detecção de doenças foram financiados pelos projetos Fapesp 2013/06884-8 e Embrapa SEG 02.14.09.001.00.00. Os resultados relacionados aos experimentos de detecção de animais foram financiados pelo projeto Fapesp 2018/12845-9. As imagens para a pesquisa em citricultura foram disponibilizadas pela PES/Fundecitrus. Além disso, as GPUs empregadas no treinamento de redes neurais foram doadas pela NVIDIA Corporation.

## 6 Referências

- BAC, C. W.; HEMMING, J.; TUIJL, B. A. J. van; BARTH, R.; WAIS, E.; HENTEN, E. J. van. Performance evaluation of a harvesting robot for sweet pepper. **Journal of Field Robotics**, v. 34, n. 6, p. 1123-1139, Sept 2017. DOI: [10.1002/rob.21709](https://doi.org/10.1002/rob.21709).
- BARBEDO, J. G. A.; KOENIGKAN, L. V.; SANTOS, T. T.; SANTOS, P. M. A study on the detection of cattle in UAV images using deep learning. **Sensors**, v. 19, n. 24, article number 5436, Dec 2019. DOI: [10.3390/s19245436](https://doi.org/10.3390/s19245436).
- BARBEDO, J. G. A. Plant disease identification from individual lesions and spots using deep learning. **Biosystems Engineering**, v. 180, p. 96-107, Apr 2019. DOI: [10.1016/j.biosystemseng.2019.02.002](https://doi.org/10.1016/j.biosystemseng.2019.02.002).
- BARBEDO, J. G. Factors influencing the use of deep learning for plant disease recognition. **Biosystems Engineering**, v. 172, p. 84-91, Aug 2018. DOI: [10.1016/j.biosystemseng.2018.05.013](https://doi.org/10.1016/j.biosystemseng.2018.05.013).
- BARGOTI, S.; UNDERWOOD, J. Deep fruit detection in orchards. In: IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 2017, Singapore. **Proceedings**. IEEE, 2017. p. 3626-3633. DOI: [10.1109/ICRA.2017.7989417](https://doi.org/10.1109/ICRA.2017.7989417).
- BIRRELL, S.; HUGHES, J.; CAI, J. Y.; IIDA, F. A field-tested robotic harvesting system for iceberg lettuce. **Journal of Field Robotics**, v. 37, n. 2, p. 225-245, Mar 2020. DOI: [10.1002/rob.21888](https://doi.org/10.1002/rob.21888).
- CAMARGO NETO, J.; TERNES, S.; SOUZA, K. X. S. de; YANO, I. H.; QUEIROS, L. R. Uso de redes neurais convolucionais para detecção de laranjas no campo. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, Indaiatuba. **Anais...** Ponta Grossa: SBIAGRO, 2019. p. 312-321. Organizadores: Maria Fernanda Moura, Jayme Garcia Arnal Barbedo, Alaine Margarete Guimarães, Valter Castelhana de Oliveira. SBIAGRO 2019. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1125722/1/PC-Redes-neurais-SBIAGRO-2019.pdf>. Acesso em: 10 out. 2020.
- COMBA, L.; BIGLIA, A.; AILONINO, D. R.; GAY, P. Unsupervised detection of vineyards by 3D point-cloud UAV photogrammetry for precision agriculture. **Computers and Electronics in Agriculture**, v. 155, p. 84-95, Dec 2018. DOI: [10.1016/j.compag.2018.10.005](https://doi.org/10.1016/j.compag.2018.10.005).
- DETONE, D.; MALISIEWICZ, T.; RABINOVICH, A. SuperPoint: self-supervised interest point detection and description. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS, 2018, Salt Lake City. **Proceedings**. IEEE, 2018. p. 337-349. DOI: [10.1109/CVPRW.2018.00060](https://doi.org/10.1109/CVPRW.2018.00060).
- DUCKETT, T.; PEARSON, S.; BLACKMORE, S.; GRIEVE, B. Agricultural robotics: the future of robotic agriculture. **UK-RAS Network**, June 2018. DOI: [10.31256/WP2018.2](https://doi.org/10.31256/WP2018.2).
- FERENTINOS, K. P. Deep learning models for plant disease detection and diagnosis. **Computers and Electronics in Agriculture**, v. 145, p. 311-318, Feb 2018. DOI: [10.1016/j.compag.2018.01.009](https://doi.org/10.1016/j.compag.2018.01.009).
- FORSMOO, J.; ANDERSON, K.; MACLEOD, C. J. A.; WILKINSON, M. E.; BRAZIER, R. Drone-based structure-from-motion photogrammetry captures grassland sward height variability. **Journal of Applied Ecology**, v. 55, n. 6, p. 2587-2599, Nov 2018. DOI: [10.1111/1365-2664.13148](https://doi.org/10.1111/1365-2664.13148).
- GENÉ-MOLA, J.; GREGORIO, E.; CHEEIN, F. A.; GUEVARA, J.; SANZ-CORTIELLA, R.; ESCOLÀ, A.; ROSELL-POLO, J. R. Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced air flow. **Computers and Electronics in Agriculture**, v. 168, 105121, Jan 2020. DOI: [10.1016/j.compag.2019.105121](https://doi.org/10.1016/j.compag.2019.105121).
- GONGAL, A.; AMATYA, S.; KARKEE, M.; ZHANG, Q.; LEWIS, K. Sensors and systems for fruit detection and localization: a review. **Computers and Electronics in Agriculture**, v. 116, p. 8-19, Aug 2015. DOI: [10.1016/j.compag.2015.05.021](https://doi.org/10.1016/j.compag.2015.05.021).

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge; London: MIT Press, 2016.

HÄNI, N.; ROY, P.; ISLER, V. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. **Journal of Field Robotics**, v. 37, n. 2, p. 263-282, Mar 2020.

HARTLEY, R.; ZISSERMAN, A. **Multiple view geometry in computer vision**. 2nd ed. New York: Cambridge University Press, 2003. DOI: [10.1017/CBO9780511811685](https://doi.org/10.1017/CBO9780511811685).

HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. Mask R-CNN. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2017, Venice. **Proceedings**. IEEE, 2017. p. 2980-2988. DOI: [10.1109/ICCV.2017.322](https://doi.org/10.1109/ICCV.2017.322).

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: a survey. **Computers and Electronics in Agriculture**, v. 147, p. 70-90, Apr 2018. DOI: [10.1016/j.compag.2018.02.016](https://doi.org/10.1016/j.compag.2018.02.016).

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, May 2015. DOI: [10.1038/nature14539](https://doi.org/10.1038/nature14539).

LIU, X.; CHEN, S. W.; LIU, C.; SHIVAKUMAR, S. S.; DAS, J.; TAYLOR, C. J.; UNDERWOOD, J.; KUMAR, V. Monocular camera based fruit counting and mapping with semantic data association. **IEEE Robotics and Automation Letters**, v. 4, n. 3, p. 2296-2303, 2019. DOI: [10.1109/LRA.2019.2901987](https://doi.org/10.1109/LRA.2019.2901987).

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91-110, Nov 2004. DOI: [10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94).

MILELLA, A.; MARANI, R.; PETITTI, A.; REINA, G. In-field high throughput grapevine phenotyping with a consumer-grade depth camera. **Computers and Electronics in Agriculture**, v. 156, p. 293-306, Jan 2019. DOI: [10.1016/j.compag.2018.11.026](https://doi.org/10.1016/j.compag.2018.11.026).

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: unified, real-time object detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016, Las Vegas. **Proceedings**. IEEE, 2016. p. 779-788. DOI: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).

REDMON, J.; FARHADI, A. YOLO v3: an incremental improvement [DB]. **arXiv preprint arXiv:1612.08242**, 2018.

REDMON, J.; FARHADI, A. YOLO9000: better, faster, stronger. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2017, Honolulu. **Proceedings**. IEEE, 2017. p. 6517-6525. DOI: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690).

ROSER, M. **Employment in agriculture**. 2013. Disponível em: <https://ourworldindata.org/employment-in-agriculture>. Acesso em: 19 maio 2020.

SA, I.; GE, Z.; DAYOUB, F.; UPCROFT, B.; PEREZ, T.; MCCOOL, C. DeepFruits: a fruit detection system using deep neural networks. **Sensors**, v. 16, n. 8, p. 1222, Aug 2016. DOI: [10.3390/s16081222](https://doi.org/10.3390/s16081222).

SANTOS, T. T.; BASSOI, L. H.; OLDONI, H.; MARTINS, R. L. Automatic grape bunch detection in vineyards based on affordable 3D phenotyping using a consumer webcam. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 11., 2017, Campinas. **Ciência de dados na era da agricultura digital**: anais. Campinas: Editora da Unicamp; Embrapa Informática Agropecuária, 2017. p. 89-98. SBIAGro 2017. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1083291/1/AutomatcgrapeSBIAGro.pdf>. Acesso em: 16 out. 2020.

SANTOS, T. T.; SOUZA, L. L. de; SANTOS, A. A. dos; AVILA, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. **Computers and Electronics in Agriculture**, v. 170, article 105247, p. 1-17, Mar 2020. DOI: [10.1016/j.compag.2020.105247](https://doi.org/10.1016/j.compag.2020.105247).

STACHNISS, C.; LEONARD, J. J.; THRUN, S. Simultaneous localization and mapping. In: SICILIANO, B.; KHATIB, O. (ed.). **Springer handbook of robotics**. Cham: Springer International Publishing, 2016. p. 1153-1176. DOI: [10.1007/978-3-319-32552-1\\_46](https://doi.org/10.1007/978-3-319-32552-1_46).

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015, Boston. **Proceedings**. IEEE, 2015. p. 1-9. DOI: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).

WESTOBY, M.; BRASINGTON, J.; GLASSERA, N. F.; HAMBREY, M. J.; REYNOLDS, J. M. 'Structure-from-Motion' photogrammetry: a low-cost, effective tool for geoscience applications. **Geomorphology**, v. 179, p. 300-314, Dec 2012. DOI: [10.1016/j.geomorph.2012.08.021](https://doi.org/10.1016/j.geomorph.2012.08.021).

WILLIAMS, H.; TING, C.; NEJATI, M.; JONES, M. H.; PENHALL, N.; LIM, J.; SEABRIGHT, M.; BELL, J.; AHN, H. S.; SCARFE, A.; DUKE, M.; MACDONALD, B. Improvements to and large-scale evaluation of a robotic kiwifruit harvester. **Journal of Field Robotics**, v. 37, n. 2, p. 187-201, Mar 2020. DOI: [10.1002/rob.21890](https://doi.org/10.1002/rob.21890).

XIONG, Y.; GE, Y.; GRIMSTAD, L.; FROM, P. J. An autonomous strawberry-harvesting robot: design, development, integration, and field evaluation. **Journal of Field Robotics**, v. 37, n. 2, p. 202-224, Mar 2020. DOI: [10.1002/rob.21889](https://doi.org/10.1002/rob.21889).

ZOPH, B.; VASUDEVAN, V.; SHLENS, J.; LE, Q. V. Learning transferable architectures for scalable image recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2018, Salt Lake City. **Proceedings**. IEEE, 2018. p. 8697-8710. DOI: [10.1109/CVPR.2018.00907](https://doi.org/10.1109/CVPR.2018.00907).