



15º Congresso Interinstitucional de Iniciação Científica – CIIC 2021
01 a 02 de setembro de 2021
ISBN 978-65-994972-0-9

PIPELINE DE DETECÇÃO DE LARANJAS A PARTIR DE VÍDEOS

Tiago Petená **Veraldi**¹; João **Camargo Neto**²; Thiago Texeira **Santos**³; Sônia **Ternes**⁴; Kleber
Xavier Sampaio De **Souza**⁵

Nº 21609

RESUMO - A detecção de frutos utilizando vídeos, adquiridos em laranjas, é um processo que envolve a utilização de vários sistemas. Cada um é responsável por uma etapa do processo de detecção, sendo que o resultado de um serve como entrada para o outro. Para que o processo seja executado corretamente é necessário a validação dos resultados em cada etapa, antes do início da próxima etapa. Caso o resultado de uma etapa não seja satisfatória, é necessário executar a etapa anterior utilizando outros parâmetros de ajustes, até que se obtenha um resultado aceitável. Este procedimento pode ser executado manualmente, mas é muito trabalhoso. Este trabalho apresenta uma solução baseada no conceito de máquina de estado finito para automatizar o processo de detecção de frutos.

Palavras-chaves: Máquina de Estado, Visão Computacional, Redes Neurais Profundas, Framework.

1 Autor, Bolsista CNPq (PIBITI): Graduação em Engenharia da Computação, Unicamp, Campinas-SP; tiagopetena@gmail.com.

2 Orientador: Analista da Embrapa Informática Agropecuária, Campinas-SP; joao.camargo@embrapa.br.

3 Pesquisador da Embrapa Informática Agropecuária, Campinas-SP.

4 Pesquisadora da Embrapa Informática Agropecuária, Campinas-SP.

5 Pesquisador da Embrapa Informática Agropecuária, Campinas-SP.



15º Congresso Interinstitucional de Iniciação Científica – CIIC 2021
01 a 02 de setembro de 2021
ISBN 978-65-994972-0-9

ABSTRACT – *Estimating fruit yield is an essential step of great importance to productive and distributive chains. Algorithms using computer vision for this end are increasingly becoming more robust and producing remarkable results, taking over traditional methods that may be inefficient. The process of running these algorithms, however, can also become very complex, requiring multiple frameworks, parameter tinkering and specialized knowledge while being time and resource dependent. This work aims to streamline the execution of an orange yield estimation algorithm making use of state machines, separating the different frameworks into different states and making automatic transitions between those, reducing both execution and distribution overhead.*

Keywords: State Machine, Computer Vision, Deep Neural Networks, Framework.

1 INTRODUÇÃO

Este trabalho está sendo desenvolvido no âmbito do projeto eContaFruto, “Estimativa da quantidade de frutos em pés de laranja por meio de inteligência computacional” (TERNES, 2019), desenvolvido e liderado pela Embrapa Informática Agropecuária, em parceria com o Fundo de Defesa da Citricultura – Fundecitrus, dentro do Programa de Estimativa de Safra (PES). O objetivo do projeto eContaFruto é a utilização de tecnologias modernas de visão computacional para estimar a quantidade de frutos por árvore antes da colheita. Numa primeira fase, o trabalho de detecção utilizou imagens estáticas para realizar a contagem dos frutos visíveis na copa da árvore. Neste processo a rede neural de aprendizado profundo YOLO-V3 (REDMON; FARHAD, 2018) foi treinada e a partir de uma estratégia de janela deslizante toda a copa da árvore foi processada



15º Congresso Interinstitucional de Iniciação Científica – CIIC 2021
01 a 02 de setembro de 2021
ISBN 978-65-994972-0-9

para a estimativa geral da quantidade de frutos visíveis. Os resultados foram satisfatórios, mas este procedimento só analisava a copa da árvore em um ângulo de visão, no qual a foto foi adquirida. Para que esta análise fosse realizada em vários ângulos de visão, proporcionando uma melhora na detecção de frutos oclusos, a utilização de vídeos era necessária. O trabalho de detecção de frutos utilizando vídeos já vem sendo utilizado com sucesso em outras culturas, como em parreiras (SANTOS et al., 2020) e maçãs (ITAKURA et al., 2021).

A estimativa do número de laranjas a partir de um vídeo envolve várias etapas, cada uma destas envolve parâmetros ajustáveis, diferentes requisitos e um longo tempo de processamento. Aplicar o processo manualmente em centenas de árvores se torna exaustivo e, portanto, praticamente inviável.

Este trabalho visa automatizar o processamento dos vídeos - fazendo ajustes de parâmetro específicos a cada vídeo e transições automáticas entre os estágios. Para isso, são construídas máquinas de estado para cada vídeo, em que cada estado corresponde a uma etapa diferente e as transições são acionadas com base nos seus resultados.

2 MATERIAL E MÉTODOS

Os vídeos dos pés de laranja utilizados neste trabalho foram adquiridos com *smartphones* em diversas regiões dos estados de São Paulo e Minas Gerais nos quais o Fundecitrus realizou o



15º Congresso Interinstitucional de Iniciação Científica – CIIC 2021
01 a 02 de setembro de 2021
ISBN 978-65-994972-0-9

trabalho de estimativa de safra do ano de 2021. Estes vídeos são capturados manualmente e cada vídeo contém apenas um lado da árvore, isso se dá pela dificuldade de transitar ao redor da planta inteira em um só vídeo.

A entrada do algoritmo de contagem dos frutos é apenas o vídeo, o que faz com que seja necessário que um QR-CODE seja mostrado à câmera no início do vídeo para que a identificação da árvore possa ser feita.

Após a identificação da árvore, o vídeo é separado em '*frames*' (quadros) - imagens estáticas que compõem o vídeo. Quanto mais *frames* forem gerados, mais longo será o processo de contagem, por isso é necessário achar o número mínimo de quadros extraídos que possibilite uma boa acurácia nos módulos subsequentes. Estes *frames* são então submetidos ao processo de reconstrução de estrutura ("*Structure from motion*") do COLMAP (SCHÖNBERGER; FRAHM, 2016). Se o número de quadros for suficiente para a reconstrução, a rede de detecção de laranjas YOLO-V3 processa cada frame. A última etapa é achar a correspondência entre a estrutura gerada pelo COLMAP e as laranjas detectadas para que seja possível fazer o rastreamento das laranjas no espaço tridimensional e reduzir duplicidades, para que uma boa estimativa seja feita.

2.1 ESTADOS

A máquina de estados (Figura 1) foi implementada em Python utilizando a biblioteca pytransitions (YARKONI et al., 2014). Esta possibilitou uma fácil definição de estados, gatilhos e



transições. O andamento do processo é salvo para que, caso este seja interrompido, seja possível retornar ao ponto anterior sem repetir todos os passos.

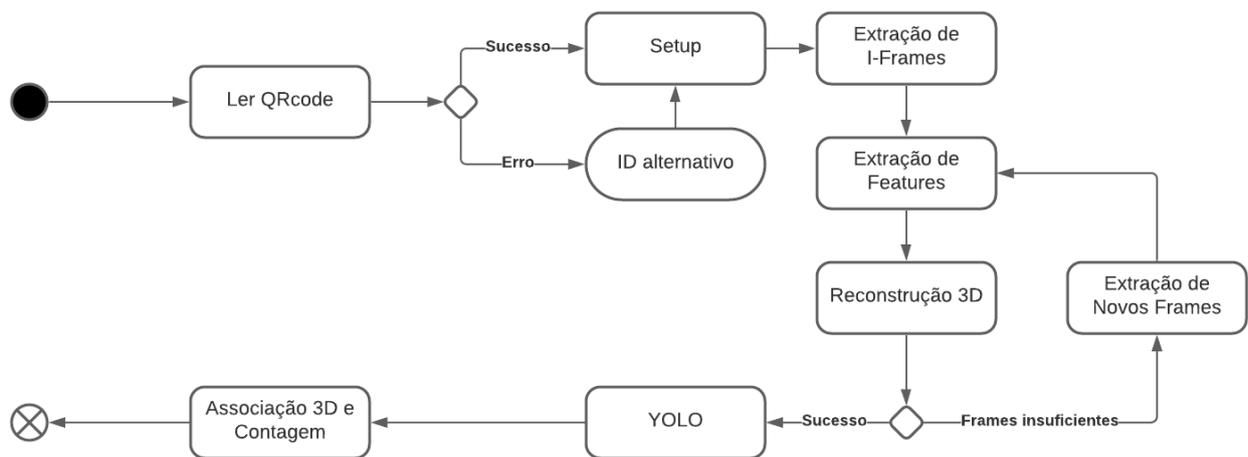


Figura 1. Diagrama simplificado da máquina de estados.

2.2 ANÁLISE DE FRAMES

A separação dos *frames* é feita utilizando o software *open-source* de manipulação de imagem e vídeo *FFmpeg*. Essa divisão pode ocorrer de duas maneiras: A primeira, e ideal, é feita selecionando os *I-Frames*(ou *intra-coded frames*) - quadros independentes uns dos outros durante a compressão/descompressão dos vídeos - entretanto, estes podem não aparecer em número suficiente para a reconstrução 3D. Neste caso, a segunda opção é utilizada: é definido uma frequência mínima de amostragem que é elevada gradualmente até que a reconstrução seja possível.



2.3 RECONSTRUÇÃO 3D

Após a obtenção dos quadros, estes são submetidos ao processo de reconstrução 3D “Structure from motion” utilizando o COLMAP, software especializado neste fim (SCHÖNBERGER; FRAHM, 2016). O algoritmo baseia-se na movimentação da câmera e em seus parâmetros intrínsecos para gerar uma nuvem de pontos 3D (*point-cloud*). Uma representação da estrutura de toda a árvore é gerada por meio de pontos em coordenada 3D.

Este processo envolve obter a posição da câmera no momento da captura da cada imagem, algo possível pela identificação de *landmarks*, regiões pontuais que são reconhecidas em múltiplas imagens. Esses *landmarks* são empregados por algoritmos de otimização não-linear para a estimação da posição tridimensional da câmera em cada quadro do vídeo e da localização 3-D dos próprios *landmarks* (HARTLEY; ZISSERMAN, 2003).

2.4 DETECÇÃO DE LARANJAS

O próximo passo é localizar as laranjas em cada *frame*. Essa localização é feita utilizando a rede neural YOLO-V3 com a técnica de janela deslizante implementada no trabalho anterior (CAMARGO et al., 2019). A sua execução é feita, preferencialmente, em placas gráficas. O *backend* da Darknet (REDMON et al., 2016) é criado com biblioteca *open-source* de visão computacional, OpenCV. O posicionamento de cada laranja, representado pelas coordenadas que a rede tem como saída, é então obtido em todos os *frames* selecionados.



2.5 ASSOCIAÇÃO 3D E CONTAGEM

Na última etapa, é feita a projeção das laranjas encontradas sobre a nuvem de pontos 3D gerada. É realizada então uma filtragem para facilitar a fusão de laranjas próximas, de modo a evitar que uma mesma laranja seja contada em localizações diferentes e, ao mesmo tempo, evitando com que laranjas diferentes sejam contadas como apenas uma.

3 RESULTADOS E DISCUSSÃO

A pipeline foi executada em diversos vídeos para uma validação inicial, e já é possível observar resultados satisfatórios no número de laranjas detectadas. Devido à não uniformidade dos métodos e condições de extração dos vídeos, ainda é necessário analisar a performance em um conjunto maior de dados.

Na Figura 2 é mostrado um exemplo de detecção dos frutos em dois *frames* sequenciais. Em amarelo os frutos já detectados em *frames* anteriores, em vermelho frutos novos detectado e em azul frutos detectados pela rede YOLO-V3 em 2D mas não detectado pelo processo 3D. Os últimos se devem a falhas na correspondência entre os quadros durante a reconstrução 3D e podem futuramente serem usados como parâmetro para ajustes em como a estrutura 3D é gerada.



Figura 2. Exemplo de sequência de *frames* no processo de detecção 3D.

Também é necessário encontrar os parâmetros ótimos de distância entre as laranjas a serem unificadas e a frequência de amostragem de *frames* a serem extraídos na máquina de estados. Um exemplo de nuvem de pontos pode ser visto na Figura 3.

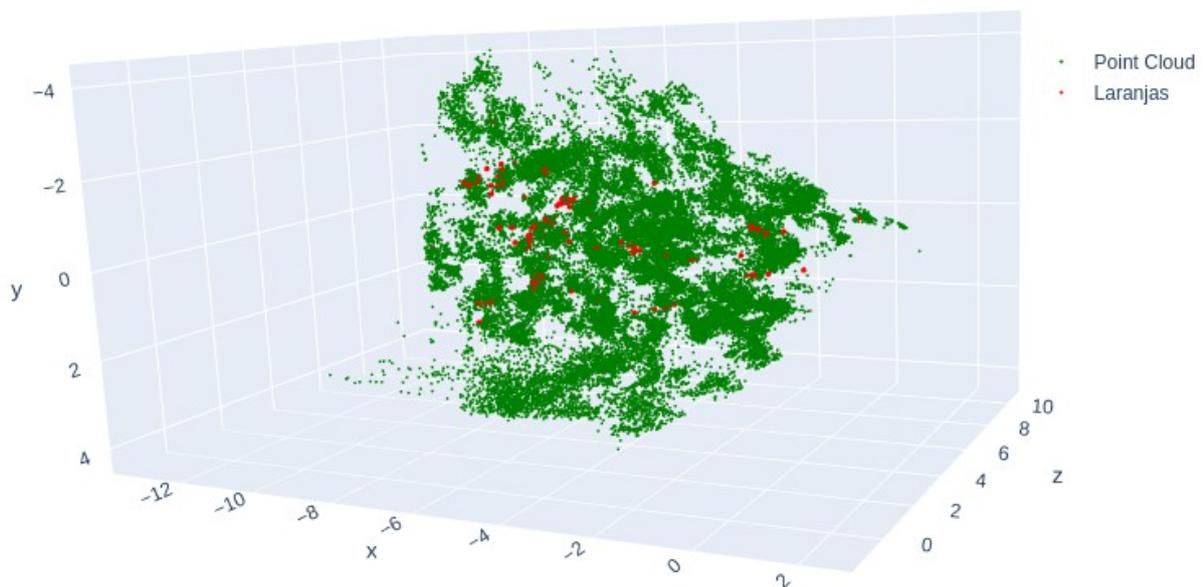


Figura 3. Point-cloud gerada com projeção das laranjas.

4 CONCLUSÃO

A trilha de detecção de laranjas possui muitas etapas e requisitos para ser reproduzida manualmente - este projeto não só facilita sua execução, mas define um guia de como os parâmetros serão escolhidos, possibilita a distribuição em diversas máquinas e melhora a visualização dos resultados.



15º Congresso Interinstitucional de Iniciação Científica – CIIC 2021
01 a 02 de setembro de 2021
ISBN 978-65-994972-0-9

No atual estágio de desenvolvimento do trabalho, já é possível fazer a contagem de laranjas automaticamente com os ajustes de parâmetros necessários. A contagem de laranja se torna mais simples já que a etapa de supervisão humana pode ser feita apenas observando os relatórios gerados em cada execução. Estes ajustes, entretanto, ainda precisam ser regulados para que seja atingido o balanço ótimo entre tempo de processamento e acurácia.

O empacotamento em um contêiner “Docker” também se mostra necessário devido à dificuldade de instalação dos requisitos para a execução das etapas, principalmente quando executado em GPU, e já está sendo implementado. O contêiner também facilitará sua distribuição e execução em sistemas alternativos.

5 AGRADECIMENTOS

Agradecemos ao PIBITI/CNPq pela bolsa concedida, a Embrapa pela oportunidade de trabalho no projeto e à equipe do PES/Fundecitrus pela disponibilização dos vídeos de pés de laranja.

6 REFERÊNCIAS

CAMARGO NETO, J.; TERNES, S.; SOUZA, K. X. S.; 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 Castelhamo de Oliveira. SBIAgro 2019. Disponível em:



15º Congresso Interinstitucional de Iniciação Científica – CIIC 2021
01 a 02 de setembro de 2021
ISBN 978-65-994972-0-9

<<https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1125722/1/PC-Redes-neurais-SBIAGRO-2019.pdf>>.
Acesso em: 15 jul. 2021.

HARTLEY, R.; ZISSERMAN, A. **Multiple view geometry in computer vision**. 2. ed. Cambridge: Cambridge University Press, 2003. 670 p.

ITAKURA, K.; NARITA, Y.; NOAKI, S.; HOSOI, F. Automatic pear and apple detection by videos using deep learning and a Kalman filter. **OSA Continuum**, v. 4, n. 5, p. 1688-1695, 2021. DOI: 10.1364/OSAC.424583.

REDMON, J. **Darknet**: Open Source Neural Networks in C. Disponível em: <<http://pjreddie.com/darknet/>>.
Acesso em: 21 jun. 2021.

REDMON, J.; FARHAD, A. **YOLOv3**: an incremental improvement. 2018. Disponível em: <<https://arxiv.org/abs/1804.02767>>. Acesso em: 21 jun. 2021

SANTOS, T. T.; SOUZA, L. L.; 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, p. 1-17, 2020. DOI: 10.1016/j.compag.2020.105247.

SCHÖNBERGER, J. L.; FRAHM, J. M. Structure-from-Motion revisited. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 29., 2016, Las Vegas. **Proceedings...** Piscataway: IEEE, 2016. p. 4104-4113.

TERNES, S. **Estimativa da quantidade de frutos em pés de laranja por meio de inteligência computacional**. Centro Nacional de Pesquisa Tecnológica em Informática para Agricultura, 2019. 25 p. (Embrapa. Tipo I – Pesquisa e Desenvolvimento – Código SEG 10.18.03.016.00.00). Projeto em andamento: eContaFruto.

YARKONI, T.; NEUMANN, A. **Pytransition**, Disponível em: <<https://github.com/pytransitions/transitions>>.
Acesso em: 28 Jun. 2021.