Universidad de Huelva

Departamento de Ingeniería Electrónica, de Sistemas Informáticos y Automática

Computer Vision in Oliviculture. Contributions to the Postharvest Estimation of Individual Fruit Features, Early In-the-field Yield Prediction, and Individual Tree Characterisation from Aerial Imagery, by means of Image Analysis

Memoria para optar al grado de doctor
presentada por:

Juan Manuel Ponce Real

Fecha de lectura: 9 de noviembre de 2020

Bajo la dirección de los doctores:

José Manuel Andújar Márquez

Arturo Aquino Martín

Huelva, 2020

Universidad de Huelva

Programa de Doctorado de Ciencia y Tecnología Industrial y Ambiental

Línea de investigación: Ingeniería Eléctrica, Electrónica, de Control y Robótica

Universidad de Huelva

Computer Vision in Oliviculture. Contributions to the Postharvest Estimation of Individual Fruit Features, Early In-the-field Yield Prediction, and Individual Tree Characterisation from Aerial Imagery, by means of Image Analysis

Memoria para optar al grado de doctor presentada por:

Juan Manuel Ponce Real

Fecha de lectura:

Bajo la dirección de los doctores:

José Manuel Andújar Márquez

Arturo Aquino Martín

Huelva, 2020

UNIVERSIDAD DE HUELVA

Programa de Doctorado de Ciencia y Tecnología Industrial y Ambiental

Línea de investigación: Ingeniería Eléctrica, Electrónica, de Control y Robótica

Visión por Computador en la Olivicultura. Contribuciones a la Estimación Postcosecha de Características Individuales de Aceitunas, a la Predicción de Producción, y a la Caracterización Individual de Olivos a partir de Imágenes Aéreas, mediante Técnicas de Análisis de Imagen

Tesis Doctoral

JUAN MANUEL PONCE REAL

2020

Directores:

Dr. José Manuel Andújar Márquez

Dr. Arturo Aquino Martín

Dr. José Manuel Andújar Márquez, Catedrático de Universidad de la Escuela Técnica Superior de Ingeniería de la Universidad de Huelva,

Dr. Arturo Aquino Martín, Investigador PostDoc de la Universidad de Huelva,

CERTIFICAN:

Que D. Juan Manuel Ponce Real, Ingeniero en Informática por la Universidad de Sevilla, ha realizado bajo nuestra dirección y dentro del programa de doctorado de Ciencia y Tecnología Industrial y ambiental (CyTIA), y en la línea de investigación de Ingeniería Eléctrica, Electrónica, de Control y Robótica, el trabajo correspondiente a su Tesis Doctoral titulada:

Visión por Computador en la Olivicultura. Contribuciones a la Estimación Postcosecha de Características Individuales de Aceitunas, a la Predicción de Producción, y a la Caracterización Individual de Olivos a partir de Imágenes Aéreas, mediante Técnicas de Análisis de Imagen

Revisado el presente trabajo, estimamos que puede ser presentado al Tribunal que ha de juzgarlo.

Y para que así conste a efectos de lo establecido por el Real Decreto 99/2011 y por la normativa Reguladora del título de Doctor de la Universidad de Huelva, autorizamos la presentación de este trabajo en la Universidad de Huelva.

Huelva, a de de 2020

Director: Dr. José Manuel Andújar Márquez

Director: Dr. Arturo Aquino Martín

Doctorando: D. Juan Manuel Ponce Real

Computer Vision in Oliviculture. Contributions to the Postharvest Estimation of Individual Fruit Features, Early In-the-field Yield Prediction, and Individual Tree Characterisation from Aerial Imagery, by means of Image Analysis

Doctoral research dissertation report presented by Juan Manuel Ponce Real to fulfil the requirements for a PhD. Doctoral degree with International Mention at the University of Huelva.

Juan Manuel Ponce Real

The present Dissertation has been performed at the Control and Robotics Research Group, Ref. TEP-192, from the University of Huelva, under the supervision of Dr. José Manuel Andújar Márquez and Dr. Arturo Aquino Martín, who approve its defense:

Director: Dr. José Manuel Andújar Márquez

Director: Dr. Arturo Aquino Martín

Huelva, , 2020

Acknowledgements

First, I would like to thank my Thesis Directors, Dr. José Manuel Andújar and Dr. Arturo Aquino. José Manuel, thank you for opening the doors to the world of research for me and for letting me be part of this incredible group, for your guidance and help, and above all for that almost blind trust you placed in me from the very first minute. Arturo, mentor, tireless battle partner, and above all, friend, I feel unable to find enough words with which to show you my gratitude. You have made this possible.

To my fellow lab mates, Ana, Borja, Daniela, Diego, Miguel and Ricardo, as well as many other members of TEP-192. Thank you for your company, help and advice, and for those moments of laughter and coffee.

And finally, to my parents, Carmen and José Antonio, and my brother José, all sacrifice, love and dedication. And of course, to my Sara, companion of adventures and misfortunes. You all are the real authors of this Thesis.

Agradecimientos

En primer lugar, quisiera dar las gracias a mis directores de Tesis, Dr. José Manuel Andújar y Dr. Arturo Aquino. José Manuel, gracias por abrirme las puertas al mundo de la investigación y dejarme formar parte de este increíble grupo; por tu orientación y ayuda, y sobre todo por esa confianza casi ciega que depositaste en mi desde el primer minuto. Arturo, mentor, compañero incansable de batalla, y sobre todo amigo, me siento incapaz de encontrar suficientes palabras con las que mostrarte mi gratitud. Tú has hecho esto posible.

A mis compañeros de laboratorio, Ana, Borja, Daniela, Diego, Miguel y Ricardo, así como otros tantos miembros del TEP-192. Gracias por vuestra compañía, ayuda y consejo, y por esos ratitos de risas y café.

Y finalmente a mis padres, Carmen y José Antonio, y mi hermano José, todo sacrificio, cariño y entrega. Y como no, a mi Sara, compañera de aventuras y desventuras. Vosotros sois los auténticos autores de esta Tesis.

Contents

Chapter 1
Introduction

Chapter 1. Introduction

At present, olive (*Olea europaea* L.) cultivation occupies an important position within the agricultural sector. The popularity of olive-based products has exploded in the last few decades, mainly due to their gastronomic qualities, and the well-proven health benefits of their consumption. As a result, olive farming has spread all over the world beyond the Mediterranean basin [1], where most of the traditional producer countries are concentrated, and the production of olive oil and table olives, as well as its related market, has grown significantly [2]. Indeed, in accordance with the International Olive Council[1] (IOC), olive oil production has doubled in the last 30 years, and it is estimated to be around 3,144,000 tonnes in the 2019/20 campaign [3]; the growth in table olive yields is even more significant, having almost tripled over the same period, with an estimated world production of 2,925,500 tonnes for the 2019-20 crop year [4].

Olive sector must face multiple challenges to deal with the given scenario. In fact, the yield increasing of horticultural commodities to satisfy a growing demand, while maintaining acceptable production costs and approaching farming activities in an environmental sustainable way is a cross-cutting problem in modern agriculture, regardless the type of crop [5,6]. Specially since population growth expectations suggest that worldwide food demand could be increased by 100% in 2050 [7].

Technology plays a key role in addressing this situation. In recent decades, the modernisation of farming activities has been developed within the paradigm of precision agriculture (PA). This discipline promotes the improvement of the overall agricultural activities by using objective measures regarding the temporal and spatial variability observed in the crops, for optimising the use and management of farming inputs [8]. Because of its data-driven nature, this agricultural system has historically benefited from Remote Sensing technologies, as well as from the advances that Computer Vision has experienced in recent years [9–12].

Indeed, the use of Computer Vision-based techniques has gained momentum within agriculture and the agro-food industry, due to their capacities for automatically extracting knowledge from previously captured data, and their subsequently applicability within PA environments. Hence, Computer Vision in agriculture has currently drawn the attention of research community, and it can be found a vast amount of publications related to the study and applicability of this technology, within the field of smart farming and the treatment and handling of horticultural products. Issues addressed are very diverse. Early yield prediction [13–15], crop mapping [16], quality inspection [17,18] or automated grading [19,20] are just some examples of topics undertaken within this subject. Thus, results that can be consulted in the literature highlight the possibilities of Computer Vision within the agriculture sector, enabling the generation of image analysis-based frameworks for enhanced crop phenotyping, or supporting the automation of both farming and post-

[1] International Olive Council's website: https:/www.internationaloliveoil.org/

harvest tasks, thus providing effective tools with a potential impact in terms of productivity and sustainability improvement.

Given this context, this Thesis, presented as a collection of high-impact published articles, comprises a research conducted during the last three years, with the objective of assessing the use of these kind of technology within the olive sector. Certainly, over the last decades important technological advances have been introduced to improve the olive growing activity. However, even today, some of these innovations generate some uncertainty in terms of profitability, and they appear insufficient to meet the challenges faced by the olive industry. Thus, there is ample room for the improvement of olive sector under the scope of precision agriculture and the new data-driven technologies. Hence, in this study, different issues related to olive growing were addressed. This, in order to devise and develop different image analysis-based methodologies, potentially integrable into real Machine Vision systems, aimed at supporting and automating different tasks related to the phenotyping of olive cultivars, as well as the postharvest handling of olive fruits.

On the basis of the above-mentioned objective, three different lines of research were addressed. Hence, investigation was initially focused on the use of Computer Vision techniques to support post-harvest tasks. Automated fruit grading is a fairly important research niche in this field [19], and the olive-manufacturing industry could benefit from systems that could carry out fruit size-categorisation in an autonomous and non-invasive way [21,22]. Therefore, it was proposed the development of methodologies for the detection of olive-fruits in digital images taken in laboratory, and the estimation of morphological features of each of them, thus enabling the possibility of automating their grading and size-based categorization. Thus, after successfully testing the feasibility of applying image analysis techniques and linear feature modelling for the purpose of detecting and estimating the mass and size of olive fruits, Deep Learning (DP) technologies were explored in an attempt to increase the capacities of the developed methodology. Indeed, different Convolutional Neural Network (CNN) architectures were tried for the purpose of generating image classifiers, capable of categorising each individual fruit-pixel region regarding olive variety. As an end result, it was achieved a comprehensive framework for accurately detecting and grading olive fruits, and for classifying them attending to their variety, with the possibility of being implemented in real environments, integrated into the conveyor belts which transport the fruits.

As a second milestone, research was aimed at developing a methodology for detecting fruits on the trees themselves, in images of entire olives directly taken in the orchards, as a first step in the development of a solution to automate the in-the-field- yield estimation by means of Computer Vision techniques. Indeed, early preharvest yield estimation is a valuable measure for farmers [23], but they have traditionally addressed such task subjectively, by visually analysing the amount of visible fruit directly in the field. That said, the investigation resulted in an Artificial-Vision algorithm able to detect visible olives in digital images of olive trees captured directly in the field, at night-time with artificial illumination, by applying CNN-based image classifiers. The procedure achieved is a promising first step in the prospective development of a comprehensive solution for olive-fruit yield estimation, in which image acquisition would be carried out by autonomous robotic platforms, currently under investigation.

Finally, the last issue addressed in the present dissertation was focused on the use of remotely sensed aerial imagery within the olive sector. Aerial imagery has been strongly linked to agriculture since early 1970's, from when the first modern works related to the use of satellite images for land cover classification date [24,25]. Notwithstanding, the popularisation of unmanned aerial vehicles (UAV) experienced in recent years has boosted the use of remote sensed aerial imagery in agricultural applications [26,27]. With the aim of assessing this technology in olive orchards, a set of experiments were conducted in an attempt to develop a methodology that would allow the identification of olives in aerial images, subsequently enabling the possibility of estimating individual tree features. Thus, this third line of investigation concluded with a novel methodology for transforming UAV-based multispectral captures into a binary representation by means of photogrammetry and morphological image analysis, individually segmenting the canopies of each tree appearing in such representations; thus enabling the estimation of dendrometric characteristics of each plant regarding its individual canopy.

1.1. Overview of the Thesis structure

This Thesis is composed of six themed chapters. Thus, the Thesis is properly introduced in Chapter 1, stablishing the context in which the related research was carried out, as well as the motivations for which it was undertaken. In addition, a list of the main scientific contributions derived from this research is proposed, giving a brief review of them, and highlighting the major innovations achieved.

Chapter 2 begins by laying out the objectives to be achieved during investigation, identifying the three main focuses of interest related to the olive farming around which this Thesis was carried out. Then, Section 2.2 is concerned with the experimental approach employed for reaching each of those initial objectives, presenting the different frameworks and methodologies adopted for that purpose.

The third chapter focuses on the materials and methods that were necessary for the execution of the conducted experiments. Therefore, in Section 3.1, those cultivation areas in which, depending on the nature of the experimentation, either image capture was performed, or fruit samples were collected for their later study in laboratory, are presented. Section 3.2 contains a brief description of the equipment used for image acquisition, while Section 3.3 focuses on the processing and analysis of those images acquired, overviewing the main related techniques approached during the research as well as the tools exploited for that purpose.

In Chapter 4 the main scientific works derived from the research are included in their entirety, highlighting the five articles published in high-impact journals within the fields of engineering and computing, and which support this Thesis. Additionally, some related contributions to national and international congresses are also included.

Finally, Chapter 5 brings together the principal findings of the conducted experiments, summarising and discussing the most significant achievements of this Thesis.

1.2. Scientific contributions of the Thesis

This Thesis is presented as a collection of publications. In this vein, six articles published in high-impact indexed journals are proposed as the core of the dissertation. They are the result of the different milestones reached during the conducted research, in accordance with the lines of argument originally raised. These works are listed below, providing information regarding their publication, and giving a brief review of them as well as their major novel contributions to the state of the art.

- **Article 1.** *Olive-Fruit Mass and Size Estimation Using Image Analysis and Feature Modeling.*

  Authors: J. M. Ponce, A. Aquino, B. Millan, J. M. Andújar.
  Journal: Sensors (ISSN: 1424-8220).
  Reference: vol. 18, no. 9, p. 2930.
  DOI: https://doi.org/10.3390/s18092930.
  Year: 2018.
  Quality index (Journal Citation Reports®, 2018): 23/84 (Q2) in the category "Analytical Chemistry", 12/26 (Q2) in "Electrochemistry" and 15/61 (Q1) in "Instruments & Instrumentation". Impact Factor of 3.031.

This work comprises the first step in the development of a comprehensive solution for the automated individual categorisation of olive fruits once they have been collected, with a potential applicability in real scenarios supporting postharvest tasks.

In this sense, the initial goal was to evaluate the viability of applying image analysis techniques and linear feature modelling for the purpose of estimating different characteristics of the fruits. Specifically, it was attempted the estimation of size and mass of individual fruits. With this aim in mind, it was designed an algorithm which, based on the image of a group of olive fruits photographed from above, is capable of providing an accurate estimate of the number of them that appear in the image, as well as their individual mass and size.

Unlike what happens with other types of crops, research based on the application of image analysis techniques for the characterization of the fruits is quite limited in the case of the olive. Indeed, as far as the authors' knowledge goes, no research can be found that approached the automation of olive-fruits grading, regarding their mass and size, by means of image processing and analysis. This, along with the promising results obtained during the development of this study, invited further exploration of this line of research.

- **Article 2**. *Automatic Counting and Individual Size and Mass Estimation of live-Fruits Through Computer Vision Techniques.*

  Authors: J. M. Ponce, A. Aquino, B. Millan, J. M. Andújar.
  Journal: IEEE Access (ISSN: 2169-3536).
  Reference: vol. 7, pp. 59451-59465.
  DOI: https://doi.org/10.1109/ACCESS.2019.2915169.
  Year: 2019.

Quality index (Journal Citation Reports®, 2019): 26/90 (Q2) in the category "Telecommunications", 35/156 (Q1) in "Computer Science, Information Systems" and 61/266 (Q1) in "Engineering, Electrical & Electronic". Impact Factor of 3.745.

Given the findings derived from the previous study, this second article is based on the research conducted for the purpose of improving the framework initially proposed for the estimation of morphological features of individual olive fruits.

Firstly, the number of varieties under study were increased from two to nine, in order to strengthen the generality of the proposed solution. In addition, a more realistic scenario was faced, aiming at processing images in which the fruits were stochastically distributed on the capturing area.

In order to meet the goal of improving the performance of the methodology despite the increased complexity of the images to be processed, an ad-hoc image acquisition chamber with a LED-based lighting system was crafted. This device was designed to facilitate the capture of images, increasing the contrast between the background and the fruits themselves, and thus favouring their subsequent segmentation.

The use of this framework for image capture, a new segmentation algorithm based on mathematical morphology, and a novel way of approaching the mass of the fruits based on pixel-weighting representation contributed to substantially improve the results obtained in the previous work.

- **Article 3**. *Olive-Fruit Variety Classification by Means of Image Processing and Convolutional Neural Networks.*

  Authors: J. M. Ponce, A. Aquino, J. M. Andújar.
  Journal: IEEE Access (ISSN: 2169-3536).
  Reference: vol. 7, pp. 147629-147641.
  DOI: https://doi.org/10.1109/ACCESS.2019.2947160.
  Year: 2019.
  Quality index (Journal Citation Reports®, 2019): 26/90 (Q2) in the category "Telecommunications", 35/156 (Q1) in "Computer Science, Information Systems" and 61/266 (Q1) in "Engineering, Electrical & Electronic". Impact Factor of 3.745.

The third work was a new step in the line of research explored in the previous studies. In this case, the aim was to increase the capacities of the developed framework to be able to estimate not only the weight and size of each fruit photographed, but also the variety to which it belongs. To that end, Deep Learning techniques were explored, so different Convolutional Neural Network architectures were tried for the purpose of generating image classifiers capable of categorising each individual fruit-pixel region regarding olive variety.

The classification models obtained proved to be highly accurate with the seven varieties considered in the study. Therefore, its integration with the procedures

proposed in the previous works results in a comprehensive framework for the automated individual characterization of olive fruits regarding their variety, mass and size, thus comprising a Machine Vision solution with potential applicability in the olive sector.

- **Article 4**. *Identification of olive fruit, in intensive olive orchards, by means of its morphological structure using convolutional neural networks.*

  Authors: A. Aquino, J. M. Ponce, J. M. Andújar.
  Journal: Computer and Electronics in Agriculture (ISSN: 0168-1699)
  Reference: vol. 176, p. 105616.
  DOI: https://doi.org/10.1016/j.compag.2020.105616.
  Year: 2020.
  Quality index (Journal Citation Reports®, 2019): 24/109 (Q1) in the category "Computer Science, Interdisciplinary Applications". Impact Factor of 3.858.

In Article 4 Deep Learning-based technology was also exploited, but in this case in a very different scope. As set out and discussed in the manuscript, early preharvest yield estimation can be a valuable statistic for farmers and producers. Hence, this study was focused on the development of a methodology for detecting olive fruits in images of the trees directly taken in the orchards, as a first step in the development of a solution to automate the in-the-field- yield estimation in olive farming by means of Machine Vision techniques.

To that end, images of olive trees, individually photographed at night-time with artificial illumination, are processed on the basis of mathematical morphology for reducing the search space, thus generating a set of subimages, each of which is potentially representative of a fruit. The detection of this set of candidates is performed by exploiting the pattern of light reflection shown by the fruits in the images taken. Then, a CNN-based classifier, trained with an ad-hoc compiled dataset called *OLIVEnet*, is used for determining whether each of these subimages/candidates contains a fruit or not. The creation and sharing of the *OLIVEnet* dataset can also be considered as a contribution of this Thesis, as it is the first dataset of these characteristics put publicly available for the scientific community, which it is expected to contribute boosting research in this field.

After testing the methodology on the basis of some of the most relevant CNN architectures at present, high accuracy classification rates were reached, reinforcing the viability of this solution as a promising starting point for the implementation of an actual framework for the automated early yield prediction in olive orchards.

- **Article 5**. *Automated Identification of Crop Tree Crowns from UAV multispectral Imagery by Means of Morphological Image Analysis.*

  Authors: R. Sarabia, A. Aquino, J. M. Ponce, G. López, J. M. Andújar.

Journal: Remote Sensing (ISSN: 0168-1699).
Reference: vol. 12, no. 5, p. 748.
DOI: https://doi.org/10.3390/rs12050748.
Year: 2020.
Quality index (Journal Citation Reports®, 2019): 9/30 (Q2) in the category "Remote Sensing". Impact Factor of 4.509.

Given the importance that aerial imagery has acquired within Precision Agriculture in recent years, one of the goals initially established when planning the research from which this Thesis is derived, was to assess the applicability of remotely sensed aerial imagery in the olive sector.

Therefore, a set of experiments were conducted in an attempt to develop a methodology that would allow the identification of olive trees in aerial images, subsequently enabling the possibility of estimating individual tree features. Article 5 resulted from this investigation, proposing a novel methodology which, starting from multispectral aerial captures, uses Structure from Motion (SfM) and other photogrammetric techniques to generate three-dimensional representations of the olive orchards under study. Then, by exploiting the heigh information contained in these representations, it is able to provide accurate segmentations of the canopy coverage regarding the ground, and to estimate location points for each plant by analysing the morphology of the connected components resulting from such segmentations.

It should be noted that, given the approach followed throughout the experimentation, based on exploiting elevation information instead of any colour feature or vegetation index to perform the plant projective cover segmentation, this methodology was actually conceived to be applicable not only with olive trees, but with any kind of crop tree cultivated with a regular planting pattern in intensive orchards.

- **Article 6**[2]. *A Methodology for the Automated Delineation of Crop Tree Crowns from UAV-based Aerial Imagery by means of Morphological Image Analysis.*

Authors: J. M. Ponce, A. Aquino, D. Tejada, J. M. Andújar.

This sixth article, yet to be published, rather extends the study carried out in the previous work, aiming at developing a mathematical morphology-based procedure for individually segmenting the crown projection area of each tree appearing in an aerial representation of an orchard, thus enabling the estimation of dendrometric characteristics of each plant regarding its individual canopy.

On the other hand, and as already mentioned, results reported in Article 6 suggested the potential applicability of the proposed methodology regardless the type of tree-crop. In this vein, the new solution here proposed was tested with olive trees,

[2] As of October 2020, in peer review process.

but also with lemon (*Citrus limon* L.) and orange trees (*Citrus sinesis* L.) in two different land plots.

Thus, the methodology achieved a remarkable performance in all case studies tested, even in those with a high number of occurrences of severe overlapping between the canopies of adjacent trees. Therefore, it comprises a comprehensive solution for plant inventorying and individual characterization of tree crops, from UAV-based remotely sensed multispectral data.

Chapter 2
Objectives and Methodology

Chapter 2. Objectives and Methodology

2.1. Objectives

This Thesis addresses the capabilities of Computer Vision-based technologies to provide solutions to the olive sector within the scope of Precision Agriculture. In a broad context, the ultimate goal behind the conducted research is the achievement of different methodologies based on the use of image analysis algorithms, aimed at automating the extraction of useful information to help farmers to optimise the management of farming inputs, as well as providing agronomists with efficient tools to support them in phenotyping tasks. To that end, different issues related to olive cultivation and olive-fruit postharvest handling were addressed during investigation, ranging from the development of solutions for automatically grading the fruits to the use of remotely sensed aerial photographs for mapping the orchards and individually characterising the olive trees in them. The specific objectives initially established for the development of this Thesis, and pursued during the whole investigation, are presented below:

- **Objective I. Development of models for automatically detecting fruits already harvested in images taken in laboratory, and for extracting information regarding their morphological features**.

  Given a set of olive fruits located on a flat and homogeneous surface in terms of colour, photographed from above with artificial illumination, it was proposed the development of algorithms that, by processing the images obtained, were capable of detecting each of the fruits and extracting relevant information regarding the mass and size of each of them. Additionally, the aimed methodology was also intended to allow the estimation of the variety to which each photographed fruit belongs.

- **Objective II. Development of a methodology for automatically detecting olive fruits in olive-tree images taken in the field.**

  In order to achieve a Machine Vision-based solution for early yield estimation in olive production, it was sought the development of a methodology, again based on image processing and analysis techniques, for the detection of those fruits present on olive trees, previously photographed in the field.

- **Objective III. Development of a comprehensive solution for the automated detection and characterisation of olives trees from aerial imagery.**

  The aim was to develop a comprehensive framework for acquiring and processing aerial images of olive-growing areas, taken with a UAV, in order to identify each of the trees in these images, thus enabling their automatic geolocation,

as well as obtaining information related to the morphological characteristics of their individual canopy cover.

2.2. Methodology

The methodologies approached during the research for the purpose of reaching those objectives just presented are overviewed below, briefly describing the specific procedures and techniques used throughout the experimentation. Reader is encouraged to consult the derived published documents, collected in Chapter 4, for a deeper study.

- **Methodology for Objective I. Development of models for automatically detecting fruits already harvested in images taken in laboratory, and for extracting information regarding their morphological features**.

  In short, the first step undertaken during the investigation was aimed at proving the use of image analysis-based techniques as a viable tool for estimating morphological features of olive fruits in an automated and non-invasive way. To that end, samples of olive fruits belonging to two different varieties were gathered. This set of fruits was intended to be representative in terms of variability with respect to the size and mass of the fruits.

  Then, the fruits were placed on a white regular surface, preventing them to be in touch, and photographed from above in laboratory, with controlled lighting and using a mirror-less RGB camera. After that, different approaches were assessed in order to achieve a proper segmentation of the fruits from the background in the images acquired. In this sense, different mathematical morphology-based artifacts and global statistical thresholding approaches were considered.

  Once an effective segmentation algorithm was obtained, representative data of the characteristics to be estimated for each fruit were extracted from the resulting binary images. Estimation models were subsequently generated by linearly correlating these data with objective measures of mass and size of each of the fruits, and tested by applying them on an external validation set.

  Attending to the promising results obtained, a second step was taken to improve the methodology. Therefore, there were introduced seven new olive-fruit varieties, and an ad-hoc image acquisition chamber, equipped with a LED-based lighting system, was designed to capture images of new fruit samples. With this new image acquisition framework, the quality of the initial captures was enhanced, thus enabling to address more complex scenarios, with olive fruits stochastically distributed over the capturing area. Meanwhile, the image analysis algorithm aimed at segmenting/binarizing initial captures was redesigned to cushion the changes made to the methodology.

  Finally, categorisation of the fruits regarding their variety was undertaken. Once initial captures were binarized, and all fruits individually segmented, the corresponding pixel regions were individually isolated in smaller subimages. This set of subimages, each of which contained the connected component corresponding

to a single fruit, was used to train different implementations of CNNs. In this sense, it should be noted that a rotation-based data augmentation was performed in order to increase this training dataset. The resulting image classifiers were tested, using an external validation set, yielding high rates of accuracy.

- **Methodology for Objective II. Development of a methodology for automatically detecting olive fruits in olive-tree images taken in the field.**

In the first instance, a framework for image acquisition was defined. With the aim of being able to control the illumination of the scene to be photographed, as well as to satisfy the operational requirements of a potential autonomous device that could in the future carry out the acquisition of images autonomously, the photographs were taken at night-time. The pictures were taken manually, using a mirror-less RGB camera and artificial illumination, individually photographing each of the olive trees considered in the study.

This work was approached from the beginning as an image classification problem. Therefore, the initial idea was to find a way to discriminate those pixel regions that could potentially correspond to fruits, within the olive-tree images. From this point onwards, it would be necessary to develop classification models that would finally determine which of these candidates really corresponded to olive-fruits. Thus, with the purpose of limiting the search space by somehow reducing the original image to a subset of pixel clusters, which could be approached as candidates to represent actual fruits, initial captures were processed by means of mathematical morphology, exploiting the pattern of light reflection occurred on the fruits' surface as a differentiating feature of them with respect to the other artefacts present in these images.

Once computed, these candidates were isolated in smaller subimages, and manually labelled according to whether they corresponded to an olive fruit or not. These labelled subimages were compiled, thus comprising a dataset called *OLIVEnet*, intended to be used for training and validation of CNNs, in order to finally obtain the image classifiers.

Using disjoint subsets of *OLIVEnet* for training and validation, implementations of five different CNN architectures were tested, for the purpose of determining the most suitable topology for the problem addressed.

- **Methodology for Objective III. Development of a comprehensive solution for the automated detection and characterisation of olive trees from aerial imagery.**

This line of investigation was aimed at assessing the use of UAV-based aerial imagery for mapping olive-tree cultivation areas. In this sense, the first undertaken task during investigation was to develop an appropriate framework for image acquisition. After considering and testing different alternatives, it was finally decided to use a quadcopter equipped with a 5-band multispectral camera, At the same time, a well-defined flight configuration was established, aimed at capturing

images with forward and lateral overlaps of 85% and 65% respectively, and a GSD of 0.05 m/pixel.

Once this framework was decided, and after image acquisition was performed photographing an olive orchard, it was undertaken the design and development of an image analysis methodology for transforming multispectral aerial captures into a binary aerial representation of this cultivation area, in which olive-tree canopy projections would appear segmented from the ground. To that end, during an initial stage of preprocessing, the multispectral captures were transformed into a unique 3D point cloud-based representation of the orchard, by using photogrammetry. Then, this data structure was interpolated, yielding a Digital Heigh Model (DHM) of the land plot under study.

Elevation information contained in the DHM, which was approached as a greyscale image, was exploited in order to segment those pixel areas belonging to tree crown projections from the background. This was accomplished by using morphological image analysis and statistical thresholding. Then, individual tree location points were subsequently estimated by analysing the morphology of the connected components resulting from the segmentation. It was proposed an ad-hoc procedure to automatically compute this estimation.

In a second iteration, two new cultivars were introduced for the purpose of assessing the capacities of the achieved methodology for detecting other crop-trees beyond olives. Therefore, image acquisition was carried out in two new cultivation areas, in which lemon and orange-trees were farmed. The methodology was tested considering these two new case studies.

Additionally, it was attempted the improvement of the segmentation algorithm. So, in this final stage, it was designed and implemented a novel procedure, based on morphological image analysis techniques, aimed at individually segmenting each tree canopy, delineating them even in those cases in which the crowns of adjacent trees appear overlapped.

Chapter 3
Materials

Chapter 3. Materials

The case studies considered in the various investigations carried out, as well as the equipment used during the execution of the experiments, and the main image analysis techniques approached are presented in the following subsections.

3.1. Case studies

As stated above, the research from which this Thesis derived addressed three different lines of investigation. With regard to the part of such research focused on achieving those goals pursued in Objective 1, it is appropriate to specify certain aspects and characteristics related to the fruits that were harvested to be photographed in the laboratory. In this respect, it should be noted that up to nine varieties were considered thorough the experimentation. Originally, only two olive-fruit varieties, *Arbequina* and *Picual*, were studied. Samples of both varieties were hand-picked in high-density olive orchards located in Lagar Oliveira da Serra (Ferreria do Alentejo, Portugal) in the DMS coordinates 8°10'28.7"W and 38°05'17.2"N, in January 2018. Once the scope of the investigation was extended, new samples were needed. In this second iteration, fruits were collected again by hand in orchards located in Gibraleón, province of Huelva (Andalusia, Spain), in the DMS coordinates 7°02'26.4"W and 37°20'25.9"N, in October 2018. In this case, samples of nine different olive varieties were gathered: *Arbequina*, *Arbosana*, *Picual*, *Ocal*, *Changlot Real*, *Verdial de Huévar*, *Lechín de Sevilla* and two experimental ones, respectively named as *967* and *1030*.

Regarding the experiments conducted to achieve potential Machine Vision-based solutions for early yield estimation, images of the olive trees were also taken in the latter mentioned orchard. In this case, the olive variety under study was *Picual*, and image acquisition was performed in September 2018 (two months prior harvesting).

Finally, the works related to the use of remotely sensed multispectral imagery for individually characterising crop-trees were developed around three different cultivation areas. On all of them, intensive row-based planting patterns are applied. The first one was located in Gibraleón, province of Huelva (Andalusia, Spain). Specifically, the land plot assessed can be found centred in the DMS coordinates 7°02'48.44"W and 37°20'39.80"N, and it has an approximate extent of 17.5ha. The second orchard considered, in which lemon-trees are farmed, is located nearby village of La Redondela, Isla Cristina, province of Huelva (Andalusia, Spain), centred in the coordinates 7°17'06.93"W and 37°14'07.24"N. In this case, an area of 1.37ha was photographed. In the same locality it can be found the last orchard under study: an orange groove centred in the coordinates 7°18'06.75"W and 37°13'49.55"N. In this case study, a subplot of 1.94ha was considered. Fig. 1 shows aerial images of each of these intensive orchards.

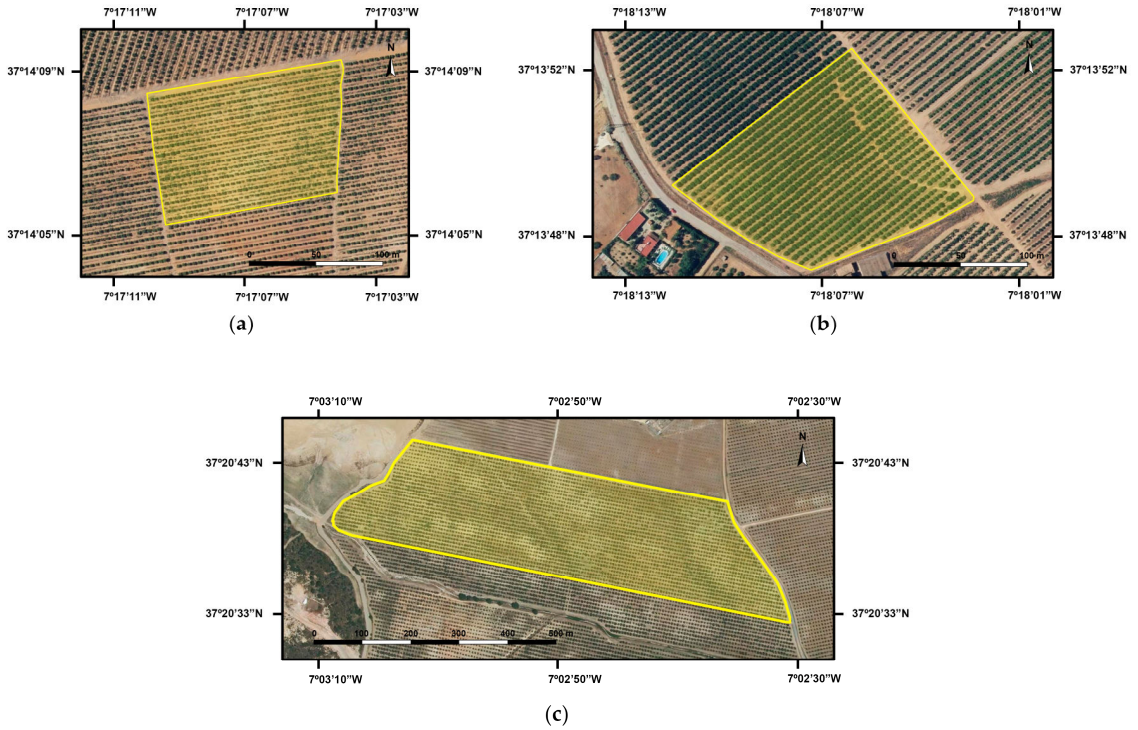


Figure 1. Aerial images of the case study sites: (a) lemon tree orchard; (b) orange tree orchard; (c) olive orchard. Note in each case, enclosed in yellow, the corresponding land plot under study.

3.2. Image acquisition equipment

Image acquisition played a key role during investigation. Indeed, given the aimed objectives, all of them focused on the development of different Computer Vision-based solutions with applicability in the olive sector, the choice of sensors and other related equipment, as well as the planning of appropriate frameworks for capturing the images, were critical issues when addressing the experimental design. Therefore, the main devices and systems used for this purpose are presented hereafter.

For photographing fruits under laboratory conditions, during the experiments carried out to develop algorithms for estimating their morphological characteristics by means of image analysis, a NMOS sensor camera was initially used; specifically, a LUMIX DMC-GH4 digital single-lens mirrorless camera (Panasonic, Kadoma, Osaka, Japan). As the research progressed, this capturing device was replaced by a Sony α7-II digital mirrorless camera (Sony Corp., Tokyo, Japan), for a second phase of data collection. It has 24 Mpx CCD stabilized sensor, and it was equipped with a Zeiss 24/70mm lens (Carl Zeiss AG, Oberkochen, Germany) with optical stabilization. In this case, this camera was installed at the top of an ad-hoc imaging chamber designed and crafted for isolating the capturing area from any external light source, and which included its own LED-based illumination system. In Fig. 2 is illustrated this image acquisition setup.

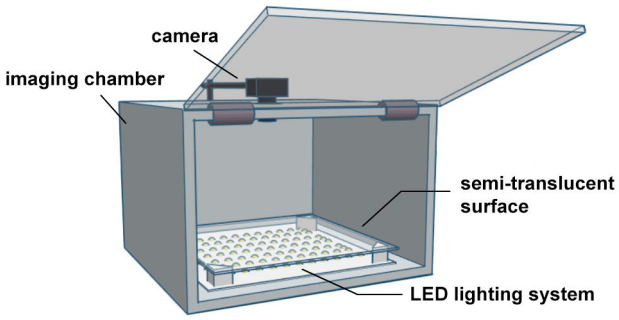
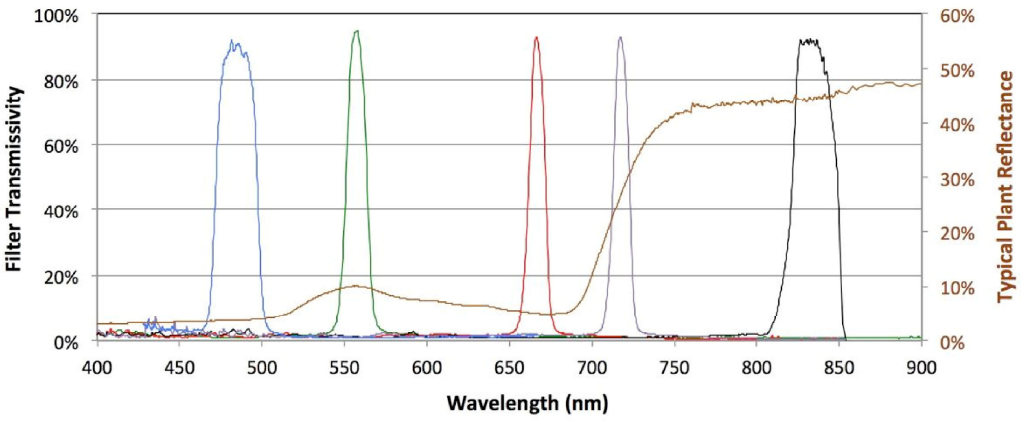


Figure 2. Image acquisition system for capturing images of olive fruits in laboratory.

The above-mentioned Sony α7-II camera was also used when photographing olive trees during execution of the research aimed at developing algorithms for detecting fruits directly in digital images of the trees.

Regarding the third line of investigation, carried out around remotely sensed aerial imagery, the aerial captures of the cultivation areas considered in the study were collected with a MicaSense RedEdge-M (MicaSense, Inc., Seattle, WA, USA) 5-band multispectral camera. This multispectral device can capture information in five different discrete spectral bands within the visible and the infrared spectrum. Its spectral sensitivity is illustrated in Fig. 3. In all cases, this camera was equipped with a dedicated GPS module, enabling georeferencing of the captured images, and a 5-band downwelling light sensor (DLS), for image correction in the event of changing light conditions. In addition, a MicaSense reflectance panel was used for determining the solar irradiance on the ground.



* Picture taken from *MicaSense RedEdge-M™ Multispectral Camera User Manual, Rev 01 – November 2017*, © 2017 MicaSense, Inc.

Figure 3. MicaSense RedEdge-M spectral response.

This multispectral camera was mounted on two different UAVs, depending on the case study. It should be remembered that the research carried out to develop image analysis procedures for the detection and characterisation of trees from aerial imagery was tested in three different orchards, in each of which is farmed a different type of crop. Thus, a DJI Matrice 100 (SZ DJI Technology Co., Ltd., Shenzhen, Guangdong, China) quadcopter was used for photographing the olive orchard. In the two remaining cases, due

to the significantly smaller extent of the plots to be studied, it could be used a UAV with lower operational range. Therefore, a DJI Phantom 3 Advance (SZ DJI Technology Co., Ltd., Shenzhen, Guangdong, China) was used for data collection. Fig. 4 shows both vehicles, together with the rest of the equipment necessary to carry out the image capture.

(**a**) (**b**)

Figure 4. Aerial image acquisition equipment: (a) DJI Matrice 100-based configuration; (b) DJI Phantom 3 Advance-based configuration.

3.3. Image processing and analysis

Being image acquisition a fundamental part in all the studies included in this Thesis, the real core of the investigation was the development of different algorithms for automatically processing the captured images, and for subsequently generating analytical models aimed at extracting meaningful information from them. The following subsections present the main techniques related to digital image analysis that were approached to devise these algorithms, as well as the necessary equipment for their implementation.

3.3.1. Image analysis techniques

3.3.1.1. Mathematical Morphology

In short, Mathematical Morphology is a theory for the processing and analysis of spatial structures, that dates back to the late sixties [28], with great applicability in digital image processing. Indeed, this theory, which is in turn based on set theory, lattice algebra and topology among others, actually comprises a set techniques that has been proven to be extremely useful when solving a huge range of image analysis-related problems [29], such as segmentation, granulometries, feature extraction, image enhancement or object recognition. Hence, morphological image analysis has been applied in countless fields, including medical imaging [30,31], materials science [32,33], industrial inspection [34], precision agriculture [15,35], geoscience [36], etc.

Mathematical Morphology-based digital image processing was originally aimed at treating binary images, approaching them as sets. But later, it was also developed for processing greyscale images, considering them as proper functions mapping a Euclidean space. In any case, the basic idea behind this technique is always the same, regardless of the type of image to be treated. Thus, morphological image analysis is rather built on the basis of transforming the image to process by probing it with a simpler structure, called structuring element (also known structuring function in greyscale morphology). This

structuring element, with a shape and size a-priori known, is transferred to the image to process, and it is applied in a similar way as convolution kernels are used in linear image filtering. So, after positioning the structuring element at all possible locations, and executing some pre-defined operation at each step, initial image is transformed. Those parts of the image that share well-defined morphological characteristics will be more or less affected depending on their affinity with the structuring element used. The way these structuring elements are applied defines a series of basic operators, from which most morphological image analysis procedures are developed. These operators are *Erosion*, *Dilation*, *Opening* ang *Closing*. Reader is encouraged to consult related literature for deeper study [29,37].

As already mentioned, Morphological Image Analysis has played a key role in the development of the works presented here. Indeed, all algorithms proposed for image processing and analysis use operators and transformations based on Mathematical Morphology; from simple morphological openings for noise filtering at postprocessing, to complex procedures based on the *Watershed* transform, for the segmentation of different artefacts. All of them can be consulted in detail in the works compiled in Chapter 4.

3.3.1.2. Convolutional neural networks

Convolutional Neural Networks are a form of Deep Learning widely used in digital image processing. Indeed, because of its well-proven capabilities when performing image classification and pattern recognition, it has rather become the state of the art in Computer Vision applications, being explored in many different fields, as robotics [38], medical imagery [39,40] or smart farming [11,41] among others.

CNNs were initially conceptualized in the eighties [42], when Kunihiko Fukushima introduced the concept of *Neocognitron* inspired by previous neuroscience studies conducted by David H. Hubel and Torsten N. Wiesel [43]. Then, this technology was actually formalised in early nineties, when the first modern work related to CNNs was published [44]. Notwithstanding, it has been in the last decade when CNNs have exploded as trendy image analysis framework, mostly due to the increase of computational capabilities offered by modern graphics processing units (GPU) [45], and the popularisation of large-scale public datasets as ImageNet [46], making available a vast amount of labelled images, thus facilitating research activity in this field.

Unlike what happens with classical neural networks, CNNs can be directly fed with raw images. Therefore, instead of preprocessing entry data by means of feature engineering for obtaining a set of descriptors with which to perform the learning process, CNN directly operates on the proper images, which is a major improvement over more classical approaches.

As a rule, a CNN architecture usually consists of two parts. The first one corresponds to a convolutional structure, aimed at extracting a set of high-level features from the entry image. This convolutional structure usually comprises sets of kernels (convolutional layers) which are convolved over the layer entry looking for characteristic patterns. In general, these convolutional layers are alternated with pooling layers for reducing the spatial dimensionality and subsequently streamlining training process, and

activation layers, usually based on rectifier linear activation functions. There is a second main structure typically implemented as a fully connected multilayer perceptron which, after conveniently trained, is the responsible for carrying out the eventual image classification. Fig. 5 proposes a schematic representation of the architecture of a generic CNN.

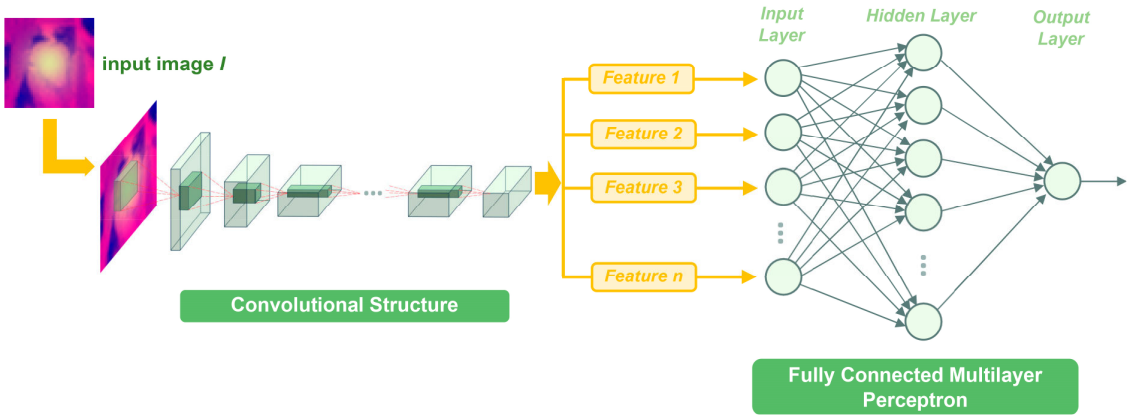


Figure 5. CNN generic architecture composed of a convolutional structure and a fully connected multilayer perceptron.

In this Thesis, CNNs have been used when addressing two of the main topics considered during the investigation conducted: the automated classification of olive-fruits regarding their variety, and the detection of them directly in images of olive-trees. In both cases, implementations of different CNN architectures were tried, in an attempt to determine the most suitable configuration for each case study. The CNN architectures approached during investigation are AlexNet [47], Inception V1 [48], Inception V3 [49], Resnet [50], Inception-ResNetV2 [51] and VGGNet [52]. Further detail is given in Chapter 4, specifically in Article 3 and Article 4.

3.3.2. Equipment for image processing and analysis

In order to implement the different image analysis algorithms and methodologies devised in the Thesis, some different software and development environments were needed. The most relevant are listed herebelow, according to the topic addressed when used:

- Codding/Implementation of image analysis algorithms, and automated procedures for testing and validation:
  - MATLAB (The Math-Works Inc., Natick, Massachusetts, USA), releases 2016 a, 2018a and 2018b.
  - MATLAB Image Processing Toolbox.
  - MATLAB Deep Learning Toolbox.
- UAV flight planning and execution:
  - DJI Flight Planner (SZ DJITM Technology Co., Ltd., Shenzhen, Guangdong, China).

  - Litchi (VC Technology, Ltd., London, UK).

- Photogrammetry:
  - Pix4Dmapper (Pix4D S.A., Prilly, Switzerland).

With regard to the hardware, it should be noted that most of the developments did not require any special equipment. However, some tasks with greater computational demand, such as the training of CNNs, did require the use of a high-performance workstation. The main characteristics of this computer are presented below:

- CPU: Intel I9 7900x (Intel Corporation, Santa Clara, CA, USA).
- GPU: Nvidia GeForce GTX 1080 (Nvidia Corporation, Santa Clara, CA, USA).
- Memory: 64GB DDR4 3300Mhz.

Chapter 4
Results

CHAPTER 4: Results

Los trabajos científicos, que forman parte del apartado Chapter 4: Results, han sido retirados de la tesis debido a restricciones relativas a derechos de autor. En sustitución de los artículos y ponencia a congreso ofrecemos la siguiente información: referencia bibliográfica, enlace al documento y resumen.

- Aquino Martín, A., Ponce Real, J.M., Millán Prior, B., Andújar Márquez, J.M.: "Post-Harvest Olive-Table Characterisation by Means of Image Analysis". En: 10th EUROSIM Congress on Modelling and Simulation. Logrono, La Rioja, Spain, July 1-5, 2019. PROCEEDINGS. (ISBN: 978-3-901608-92-6).

Enlace al texto complete del congreso:
https://www.eurosim.info/fileadmin/user_upload_argesim/ARGESIM_Publications_OA/Reports/arep.58.pdf

RESUMEN:
Post-harvest table-olive classification according to size and mass is especially important when pursuing high-quality outcomes. This paper presents a new methodology aimed at supporting accurate automatic olive-fruit grading by using artificial vision. To this end, 3,600 olive-fruit samples from nine varieties were imaged using an ad-hoc designed capturing chamber; the individuals were randomly distributed on scene. Then, an image analysis algorithm, based on mathematical morphology, was designed to individually segment olives and extract descriptive features to estimate their major and minor axes, and their mass. Next, by linearly correlating the data obtained by image analysis and the corresponding reference measurements, models for estimating the three features were computed. Then, the models were tested on 2,700 external validation samples, giving relative errors below 0.80% and 1.05% for the estimation of the major and minor axis length for all varieties, respectively. In the case of estimating olive-fruit mass, the models provided relative errors never exceeding 1.01%. The ability of the developed algorithm to individually segment olives stochastically positioned, along with the low error rates of around 1% reported by the estimation models for the three features, makes the methodology a promising alternative to be integrated in a new generation of improved and noninvasive olive classification machines.

- Aquino Martín, A., Ponce Real, J.M., Andújar Márquez, J.M.: "Identification of olive fruit, in intensive olive orchards, by means of its morphological structure using convolutional neural networks". Computers and Electronics in Agriculture. Volume 176, September 2020, 105616. https://doi.org/10.1016/j.compag.2020.105616

Enlace al texto complete del artículo: https://doi.org/10.1016/j.compag.2020.105616

RESUMEN:

Accurate yield estimation is a greatly desired objective in oliviculture due to the high economic value of its production. This paper presents a methodology aimed at achieving that end. It comprises an artificial-vision algorithm able to detect visible olives in digital images of olive trees captured directly in the field, at night-time and with artificial illumination. These images were taken in an intensive olive orchard of the Picual *Olea europaea* L. variety in September 2018 (two months prior to harvesting). Regarding the methodology, first, the images are pre-processed to generate a set of sub-images with high probability of containing an olive, thus reducing the search space by a magnitude of 103. Next, these sub-images are classified by a convolutional neural network (CNN) as *olive*, if they are centred in an olive fruit, or as *other* in any other case (even if they contain peripheral fruits). To train and validate the CNN, a special database called OLIVEnet was compiled with two disjoint sets integrating these sub-images. A training and a validation set was built with 234,168 and 299,946 *olive* and *other* sub-images, respectively. Five different CNN topologies were tested, correctly classifying the best performing one in 83.13% of *olive* instances, with a precision of 84.80%, and 99.12% of *other* instances; measured accuracy and *F1* Score were 0.9822 and 0.8396, respectively. As far as the authors' knowledge goes, this article presents the first image analysis approach to automatically identify olive fruits in an image of the entire tree directly taken in the field. The obtained results constitute a first and solid step towards the implementation of an automatic system for yield estimation of olive orchards.

- Ponce Real, J.M., Aquino Martín, A., Tejada Guzmán, D.: “A Methodology for the Automated Delineation of Crop Tree Crowns from UAV-based Aerial Imagery by means of Morphological Image Analysis”. (As of October 2020, in peer review process).

Abstract: The ability to estimate different crop field parameters directly from aerial images is revolutionising modern farming management practises. The popularisation of new forms of aerial remote sensing, mostly based on unmanned aerial vehicles (UAV), has boosted the capacities of agronomists and researchers to offer farmers valuable data regarding the status of their crops. In this paper, a novel methodology is proposed for the automated detection and individual delineation of tree crowns in aerial representations of crop fields, providing accurate information about plant population and canopy coverage in intensive-farming orchards with a row-based plant arrangement. To that end, and after pre-processing initial aerial multispectral captures by means of photogrammetry and morphological image analysis, a resulting binary representation of the land plot surveyed, where tree canopy coverage is presumably segmented from the ground, is treated at connected component-level in order to separate overlapping tree crown projections. An algorithm, also based in morphological analysis, transforms each of those components in a set of seeds with which, via *Watershed* segmentation, tree crowns are finally delineated, establishing the boundaries between them when they appear overlapped. This solution has been tested in three different orchards, in each of which is farmed a different type of fruit tree (olives, orange and lemon trees, respectively), achieving semantic segmentations with F_{score} values always above 90%, when analysing the results at pixel level. Regarding individual plant detection, trees were identified with a *precision* above 99% in all cases, and F_{score} values up to 99.79%. Attending to the results achieved, the developed methodology comprises a promising tool for automating the inventorying of plants and estimating individual tree-canopy coverage in intensive tree-based orchards, regardless of the crop farmed.

Chapter 4. Results

In the following section, the five articles published in high-impact journals that comprise this Thesis are compiled, organised according to the topic addressed. It is also included a final work yet to be published, as well as some related studies which have been presented in different congresses and symposia.

4.1. Postharvest individual fruit characterisation

4.1.1. Article 1

Olive-Fruit Mass and Size Estimation Using Image Analysis and Feature Modeling

J. M. Ponce, A. Aquino, B. Millan, J. M. Andújar

Published in:

Journal: Sensors (ISSN: 1424-8220)

Editorial: MDPI

Reference: vol. 18, no. 9, p. 2930

Year: 2018

DOI: 10.3390/s18092930

Quality index (Journal Citation Reports®, 2018): 23/84 (Q2) in the category "Analytical Chemistry", 12/26 (Q2) in "Electrochemistry" and 15/61 (Q1) in "Instruments & Instrumentation". Impact Factor of 3.031.

sensors

MDPI

Article

Olive-Fruit Mass and Size Estimation Using Image Analysis and Feature Modeling

Juan Manuel Ponce *, Arturo Aquino, Borja Millán and José Manuel Andújar

University of Huelva, Department of Electronic Engineering, Computer Systems and Automation, La Rábida, Palos de la Frontera, 21819 Huelva, Spain; arturo.aquino@diesia.uhu.es (A.A.); borja.millan@diesia.uhu.es (B.M.); andujar@diesia.uhu.es (J.M.A.)

* Correspondence: jmponce.real@diesia.uhu.es; Tel.: +34-959-217-380

Received: 30 July 2018; Accepted: 31 August 2018; Published: 3 September 2018

Abstract: This paper presents a new methodology for the estimation of olive-fruit mass and size, characterized by its major and minor axis length, by using image analysis techniques. First, different sets of olives from the varieties Picual and Arbequina were photographed in the laboratory. An original algorithm based on mathematical morphology and statistical thresholding was developed for segmenting the acquired images. The estimation models for the three targeted features, specifically for each variety, were established by linearly correlating the information extracted from the segmentations to objective reference measurement. The performance of the models was evaluated on external validation sets, giving relative errors of 0.86% for the major axis, 0.09% for the minor axis and 0.78% for mass in the case of the Arbequina variety; analogously, relative errors of 0.03%, 0.29% and 2.39% were annotated for Picual. Additionally, global feature estimation models, applicable to both varieties, were also tried, providing comparable or even better performance than the variety-specific ones. Attending to the achieved accuracy, it can be concluded that the proposed method represents a first step in the development of a low-cost, automated and non-invasive system for olive-fruit characterization in industrial processing chains.

Keywords: olive; food industry; fruit grading; image analysis; segmentation

1. Introduction

Olive growing is a high relevance agricultural activity. With a huge presence in the Mediterranean Basin, where its importance transcends the farming scope to become an actual symbol of its culture and tradition, the olive crop has spread all over the world [1,2]. Because of the well-proved health benefits of olive-derived products, and the excellence of its culinary uses, its consumption has considerably risen in recent years. According to IOC (International Olive Council) [3], table olives consumption has been increased by 173% in the twenty-five years between 1990/91 and 2015/16. Moreover, according to IOC and USDA (United States Department of Agriculture) expectations [3,4], olive oil consumption will exceed 3,000,000 tons in 2017/18.

To meet such demand, the olive industry must face multiple challenges. Despite the numbers of its market, olive farming and processing are still mainly performed in a traditional way. Even in Spain, the world largest producer, olive farming is still strongly linked to traditional production systems and low-density olive groves [5]. This model represents a problem in terms of productivity and profitability. In recent years, super-high-density olive groves, along with increased mechanization, have been introduced as response. Although some indicators suggest that these solutions, based on intensification, can provide the key for economic survival, accurate knowledge about its impact and viability is still yet to be obtained [6]. Be that as it may, within this context, the enhancement and modernization of

the processes, and the introduction of innovative solutions at all levels, are fundamental tasks to be accomplished by this industry.

Fruit sizing is a high-relevance post-harvest task in the food industry [7]. Sorting fruits and vegetables according to different attributes such as color, mass, size or shape can all be determinants in the quality and pricing of the eventual product. In the olive sector, this is especially relevant [8,9]. For table olives, the uniform size, spotless surface or appropriate coloring are determining quality-features as perceived by the final consumer. On the other hand, focusing on olive oil, size and mass of the fruits are used to calculate yield estimations. In any case, the measurement of these parameters is a necessity. However, measuring the whole harvested batch by hand is not an option due to the huge workload involved. Thus, the actual processes to extract this information are based on the study of samples.

In recent years, machine vision techniques have been explored as a valuable tool in food industry. Within the precision agriculture and the horticultural product manufacturing scopes, there is considerable literature regarding the use of image analysis to approach different problems, such as yield estimation, fruit detection, data extraction, sorting and classification or sizing and grading. Thus, Aquino et al. [10] presented a classification-based algorithm to predict grape yield at early stages from images taken on-the-go directly in the vineyard. Mery et al. [11] proposed a methodology for the detection, via image segmentation, of different kinds of food previously photographed. Cervantes et al. [12] developed a comparative analysis of different methods of feature extraction and classification of plant leaves using image processing techniques. Zhang et al. [13] developed an automatic fruit recognition system based on a split-and-merge algorithm and multiclass support vector machine (SVM). Sa'ad et al. [14] estimated the mass of mangoes from different photographs of the fruits using thresholding segmentation and provided a classification methodology supported by the information extracted from the results of this segmentation. In the same vein, Mizushima et al. [15] proposed a method for sorting and grading apples, based on the Otsu's method and linear SVM. Omid et al. [16] performed the estimation of volume and mass of citrus fruits through the segmentation of images captured in the laboratory.

Regarding the olive sector, literature focused on the application of machine vision for olive treatment and manufacturing can be consulted. Of special relevance are developments for the classification of fruits according to different characteristics, such as defects on the surface [17] or the variety [18,19], and fruit detection for feature estimation [20].

This paper proposes an efficient methodology to estimate the maximum/minimum (polar/equatorial) diameter length and mass of olive fruits by means of image analysis. To this end, as a first step, the contrast between the olives and the background is maximized in the images by employing specialized morphological processing. Then, the olives are segmented by automated thresholding based on statistical bimodal analysis. Finally, estimation models for the targeted features are obtained by correlating measurements taken from the segmentations to actual values measured in the laboratory.

The manuscript is structured as follows: Throughout three subsections, Section 2 describes the experimental design and the data acquisition process, the developed image analysis algorithm and model training for olive characterization. The next section presents the methodology proposed for result evaluation and discusses the achieved results; they have been placed together in order to provide the best understanding of the paper's research. Finally, the manuscript ends with the main conclusions on the carried-out research.

2. Materials and Methods

2.1. Reference Data and Image Acquisition

Olive fruits from two different varieties were considered for this study: Arbequina and Picual. Samples of both varieties were manually collected in January 2018, in high-density olive orchards located in Lagar Oliveira da Serra (Ferreria do Alentejo, Portugal).

Two populations (one per variety) of 200 olive fruits were selected from the samples previously acquired. Then, they were separated into different groups. Hence, for the Arbequina variety, the following sets were established: A1 (40 fruits), A2 (40 fruits), A3 (40 fruits), A4 (50 fruits) and A5 (30 fruits). For the Picual variety, four groups of 50 olives each were set up and were named as follows: P1, P2, P3 and P4.

Every described set was photographed in the laboratory, spatially distributing olives over a white plastic mat. This durable and deformable material was chosen in an attempt to approximate the type that would be used in a real conveyor belt. For capturing, the LUMIX DMC-GH4 digital single-lens mirrorless camera, equipped with a NMOS sensor, was used (Panasonic, Kadoma, Osaka, Japan). It was set up in manual mode, with an aperture of f/8, an exposure time of 1/500 s, an ISO value of 400 and a focal length of 14 mm. To reproduce an environment close to an actual industrial system, an artificial lightning setup composed of two 500 W halogen floodlights, with a light appearance of 3300 k, was employed for scene illumination. The camera was perpendicularly located above the scene; the lights were placed at the same plane and oriented to the point the camera was focused on. Figure 1 shows an example of the captured images, which were acquired and saved in JPG format, with 4608 $\times$ 2592 pixels in resolution, a pixel density of 180 ppi and a color depth of 24 bits.

Figure 1. Example of image captured of the A1 set.

To evaluate the error produced by the estimation models, objective measurements of the major and minor axis length (in millimetres—mm), and mass (in grams—g), were taken for every photographed olive by using:

- a KERN PCB 3500-2 precision balance (KERN & Sohn GmbH, Balingen, Germany).
- a 0.01 mm-resolution 0.02 mm-accuracy Electronic Digital Vernier Caliper.

The values were annotated and associated to the position of the corresponding olive fruit in the image in which it appeared.

2.2. Image Analysis and Segmentation

The proposed methodology is aimed at automatically extracting from the images features descriptive of the mass and size of the olive fruits. To accomplish this task, the developed algorithm

uses techniques based on mathematical morphology and segmentation by clustering-based image thresholding. This algorithm was implemented using MATLAB and Image Processing Toolbox Released 2016a (The MathWorks, Inc., Natick, MA, USA).

2.2.1. Preprocessing

Firstly, images are down-scaled to 40% of its original size using bicubic interpolation for the decreasing of the computational workload. Next, a *salt-and-pepper* noise reduction is accomplished by applying a gaussian filter (rotationally symmetric gaussian low-pass filter) with a standard deviation of 0.8, and a kernel size of 5×5.

Secondly, images are transformed from the native RGB color space to HSV [21]. After studying the characteristics of the images, it was concluded that the RGB space did not offer an optimal data representation for the purposes of this study. In terms of color, an absence of homogeneity between the olive fruits was detected (especially for the Arbequina variety), which prevented it from being exploited as a distinctive feature. Conversely, the difference between the fruits and the white background in terms of lightness/brightness is remarkable. The level of lightness/darkness of the color of a pixel can be accessed by transforming its RGB values in accordance with a different representation of this color model. Notwithstanding this, basing the process exclusively on light intensity could not yield good segmentation results. Indeed, there were background pixels with lightness values similar to those of olives due to the shadows cast by these fruits. At this point, it was observed that color saturation also provided object differentiation while keeping similar values for background pixels, including both the ones which belonged to a shadow and the ones that did not. Nevertheless, despite this being a partial solution to the shadow problem, the segmentation based merely on saturation couldn't yield reliable results, leading to olive pixels with saturation levels close to the background values which lacked accuracy. Therefore, neither color saturation nor intensity were found to be fully effective for image segmentation by themselves; however, an accurate combination and processing of both appeared more effective. Due to these reasons, HSV color space provided a solution, as it provides the saturation and value (level of lightness/darkness of the color) information separated into different layers (S and V channels, respectively). It is important to note that other existing color spaces are potentially valid according to this scenario, such as HSL [21] or CIELAB [22], among others.

2.2.2. Image Segmentation

Once the image is transformed into the HSV color space, the value and saturation channels are isolated into different matrices, V and S, respectively. These matrices are transformed and combined into a unique component that it is treated as a grayscale image, which is the one to be segmented. According to this, as a first step, the elements of the V component are inverted with regard to the maximum possible grey-value, i.e., 255 (for 8-bit per channel image quantification). As such, given V is the image defined in the interval [0, 255], the image V_{INV} is the one resulting from the next operation, as can be examined in Figure 2a:

$$V_{INV} = 255 - V \tag{1}$$

Considering the V channel as a greyscale image, the aim of this transformation is to set the higher grey values to olive pixels and, consequently, the lower values to the background, which becomes the darkest part of the image. Then, the saturation layer (Figure 2b) is combined to the outcome of this transformation, as is shown in Figure 2c, looking forward to improving the contrast between the background and foreground and to complement information from both sources:

$$I_{SV} = S + V_{INV} \tag{2}$$

Next, with the purpose of obtaining a background estimation, a morphological opening is applied to I_{SV}:

$$I_\gamma = \gamma_\beta(I_{SV}) = \delta_\beta\big(\varepsilon_\beta(I_{SV})\big), \tag{3}$$

where β is a 50-pixel-radius disk-shaped structuring element, large enough to contain any olive, and δ and ε are the basic morphological operations of dilation and erosion, respectively [23]. The result of this operation can be checked in Figure 2d. Then, the values of the background estimation are subtracted from I_{SV}, thus computing a high-contrast image:

$$I_{HC} = I_{SV} - I_{\gamma} \tag{4}$$

The outcome of this operation, I_{HC}, is the grayscale image to be segmented by binarization. To automatically set an optimum global threshold, the clustering-based method proposed by *Otsu* [24] was selected. This method starts from the premise that the image contains two normal-like distributions of pixels, corresponding to the foreground and the background. Then, the threshold is decided as that which maximizes the inter-class- or minimizes the intra-class-variance to optimize separation. This approach explodes the characteristics of image I_{HC}, which is the result of an image processing aimed at strengthening contrast between the olives and the background, and at homogenizing the latter to favor binarization using a global threshold. Therefore, by applying the Otsu's method to I_{HC}, the threshold *thresh* is obtained and applied to undertake its binarization as:

$$I_{BIN}(x,y) = \begin{cases} 255 \; if \; I_{HC}(x,y) > thresh \\ 0 \; in \; any \; other \; case \end{cases}, \tag{5}$$

The result of the described methodology for olive fruit segmentation can be analyzed in Figures 2 and 3.

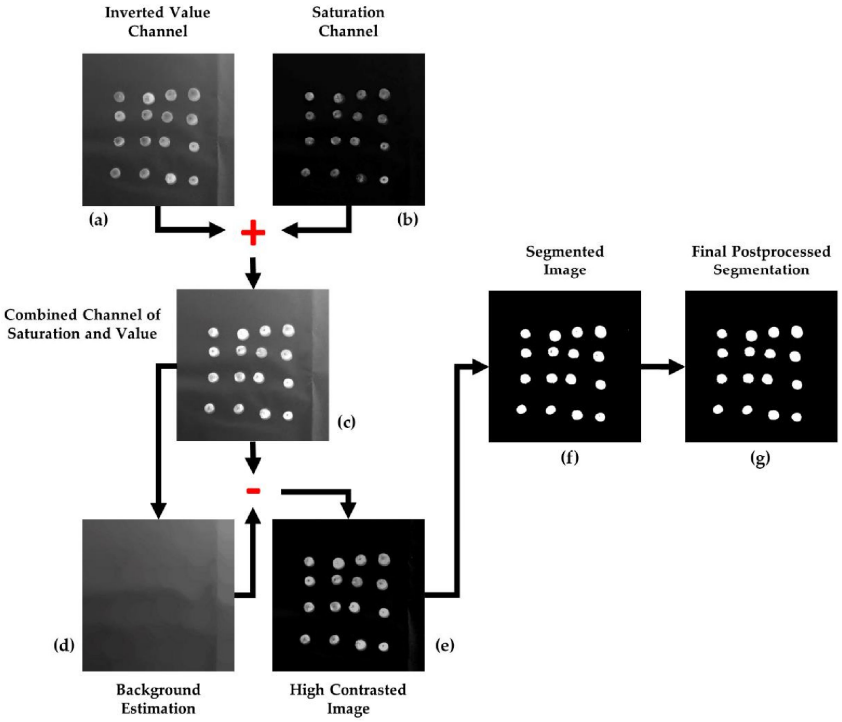


Figure 2. Step by step illustration of the image analysis algorithm on a sub-image of the study set: (**a**) inverted value channel; (**b**) saturation channel; (**c**) combined channel of saturation and value; (**d**) background estimation; (**e**) high contrasted image; (**f**) segmented image; (**g**) final postprocessed segmentation.

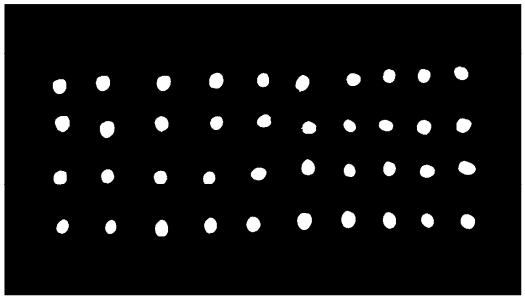


Figure 3. Result of the segmentation of the image of the A1 set, originally shown in Figure 1.

2.2.3. Postprocessing

As a last step, some morphological transformations are appealed to improve the final segmentation result. First, false positives filtering is addressed by eliminating those connected components that are too abnormally small to be considered as olive fruits. Mathematically:

$$I_{BIN2} = \gamma_\beta(I_{BIN}), \tag{6}$$

where γ is the morphological opening with a disk-shaped structuring element β with a radius of 3 pixels.

Finally, a flood-fill operation is applied to eliminate false negatives represented by the small holes which have *emerged* inside some fruit-corresponding connected components (the holes derive from points of maximum reflection of light, because of the convex surface of the fruits).

$$I_{DEF} = R^{\varepsilon}_{I_{BIN2}}(I_m), \tag{7}$$

where R is the morphological reconstruction operation, which consists on the iterative erosion (ε) of the image I_m regarding to I_{BIN2}, using a unitary structuring element, until idempotence:

$$R^{\varepsilon}_{I_{BIN2}}(I_m) = \varepsilon^{i}_{I_{BIN2}}(I_m)$$

where

$$I_m(x,y) = \begin{cases} I_{BIN2}(x,y),\ if\ (x,y)\ lies\ on\ the\ border\ of\ I_{BIN2} \\ \max(I_{BIN2}),\ otherwise \end{cases}, \text{ and} \tag{8}$$

$$\varepsilon^{i}_{I_{BIN2}}(I_m) = \varepsilon^{i+1}_{I_{BIN2}}(I_m),\ \varepsilon^{1}_{I_{BIN2}}(I_m) = \varepsilon_{\beta=1}(I_m) \vee I_{BIN2}$$

The corrective effect of this postprocessing is shown in Figure 2g.

2.3. Estimation Model Training

The goal here is to extract descriptive data from the segmented images to build estimation models for olive major and minor axis length, and mass. To this end, the binarized images allow us to work with the connected components representing the different olive fruits. First, to characterize the minor and major axis of the olives, for every component, the ellipse has the same normalized second central moments as it is being computed. Using this method, the major and minor olive axis are approximated to the major and minor axis of this ellipse, respectively, and their length in pixels is used for size estimation. On the other hand, the area of the segmented connected components, calculated as the number of constituent pixels (using 8-connectivity), is selected as a feature to estimate olive mass.

Once this information is extracted, for each of the two considered varieties, a population of 50 individuals/olives is selected as the training set; the remaining 150 individuals are kept for

external validation. These training sets are representative of the variability of the samples regarding the features under study. Next, the measurements of the major and minor axis length, and mass, corresponding to these populations and extracted automatically as specified above from the segmented images, are compared to the objective measurements taken in the laboratory. Thus, via regression analysis, linear estimation models for the targeted magnitudes and specific to each variety are yielded. Additionally, variety independent models for the magnitudes are also calculated by joining the two training sets from the two varieties and applying the same described procedure.

3. Results and Discussion

3.1. Evaluation of the Image Analysis Algorithm

Every segmented image obtained with the proposed methodology is compared to a corresponding reference image at a pixel level to evaluate its quality. To enable this comparison, a ground-truth image was generated per each image-set considered in the experiment (A1–A5 and P1–P4) by manually labeling pixels using a graphic editor (concretely Photoshop CC V 14.0, Adobe Systems Incorporated, San Jose-California, EEUU); olive and background pixel values were set to 255 and 0, respectively. Then, results of pixel comparisons are categorized and annotated according to the following definitions (see Figure 4 to check each case):

- TP: Those foreground/olive pixels in the segmented image (white pixels) matching with their analogue ones in the corresponding ground-truth image (they keep being white pixels).
- FP: Those foreground/olive pixels in the segmented image (white pixels) that were labeled as background (black pixels) in the corresponding ground-truth image.
- FN: Those background pixels in the segmented image (white pixels) that were labeled as foreground/olive (white pixels) in the corresponding ground-truth image.

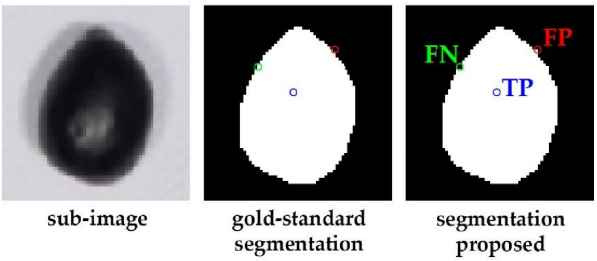


Figure 4. Examples of the different categories of pixels established to evaluate the segmentation results: *true positives* (TP, in blue); *false positives* (FP, in red); *false negatives* (FN, in green).

Thereby, segmentation quality can be finally assessed using the widely used Precision (*PC*) and Recall (*RC*) metrics, which are formulated as:

$$PC = \frac{TP}{TP + FP} \tag{9}$$

$$RC = \frac{TP}{TP + FN} \tag{10}$$

Thus, *PC* calculates the rate of correctly-detected olive pixels, and *RC* gives the rate of the actual olive. Finally, as a metric combining both *PC* and *RC* to provide with an overall accuracy measure of the segmentation method, *F*-score was calculated using the next common definition:

$$F = 2\frac{PR \times RC}{PR + RC} \tag{11}$$

3.2. Results of the Image Analysis Algorithm

As was mentioned previously, the validity of the image-segmentation algorithm has been tested through ground-truth image comparison. Based on the measures proposed to evaluate the algorithm performance, the yielded results are shown in Table 1.

Table 1. Performance of the segmentation algorithm calculated by comparison between the binary images automatically produced and the corresponding ground truths. Results are expressed in terms of Recall (RC), Precision (PC) and F-score, and detailed per variety, subset, and considering all the samples as a whole.

Set/Image	RC	PC	F-Score
Arbequina			
A1	0.9614	0.9372	0.9491
A2	0.9545	0.9432	0.9488
A3	0.9551	0.9535	0.9543
A4	0.9536	0.9745	0.9639
A5	0.9510	0.9510	0.9510
Overall	0.9551	0.9519	0.9534
Picual			
P1	0.9464	0.9810	0.9634
P2	0.9414	0.9922	0.9661
P3	0.9380	0.9869	0.9618
P4	0.9316	0.9967	0.9631
Overall	0.9393	0.9892	0.9636
Arbequina + Picual			
Overall	0.9481	0.9685	0.9580

Generally speaking, there are no outstanding differences between the two varieties in terms of algorithm performance. This fact suggests the method's viability as a variety-independent method, and it supports the initial decision of not basing it on color features (due to hue usually being a differential distinctive feature among varieties).

Interestingly, it is noticeable that high PC values were obtained, despite the lack of uniformity of the background (due to the folds of the plastic mat, as can be observed in Figure 1). This lack of uniformity implies more noise, which could provoke the increase of false positive pixels (FP), and, consequently, the impoverishment of the results in terms of precision. To avoid this phenomenon, the estimation and subtraction of the background are performed. For the hypothetical implementation of the presented method in an actual system, other materials could be explored in order to obtain a more homogeneous background, thus favoring the method's reliability. Nevertheless, since it can't be expected in an ideal scenario, background estimation/subtraction must be considered as a key part of the method.

On the other hand, it is important to underscore that the number of connected components isolated by the methodology was exactly matched with the number of olive fruits in all the images. This result is especially remarkable when considering a future commercial application of the presented solution for counting olives in a processing chain.

3.3. Evaluation of the Estimation Models

To evaluate the quality of the estimations produced by the different estimation models on the external valuation sets, the following metrics are proposed:

- Root-Mean-Square Error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (12)$$

- Relative Root-Mean-Square Error expressed as percentage

$$SE = \frac{RMSE}{\frac{\sum_{i=1}^{n} y_i}{n}} \times 100 \tag{13}$$

- Relative Mean Error expressed as percentage

$$|E| = \frac{|\sum_{i=1}^{n} (\hat{y}_i - y_i)|}{\sum_{i=1}^{n} y_i} \times 100 \tag{14}$$

where, for a feature under study (major axis, minor axis or pixel-area/mass), $\hat{y}_i$ is the predicted value and y_i is the actual value (measured previously in the laboratory), for the i-th olive-fruit processed; n is the total number of olive fruits considered.

In addition, one-way analysis of variance was addressed on the estimation results of the different developed models for the two varieties. Mean comparison was attempted, using the Tukey's test [25] at $p < 0.05$, on the population of individual relative errors defined as the ratio between the estimated and the actual value considered:

$$e_i = \frac{\hat{y}_i}{y_i}, \tag{15}$$

where $\hat{y}_i$ and y_i has the same meaning as defined above.

3.4. Results of the Estimation of Olive Features

First, the results of the correlation study performed on the training sets to obtain the estimation models are analyzed here. Separately for the training sets of Arbequina and Picual, the correlations found between the series of data pairs, 'magnitude measured in image' vs. 'actual magnitude measured in the laboratory', for the features under study, are illustrated in Figures 5 and 6.

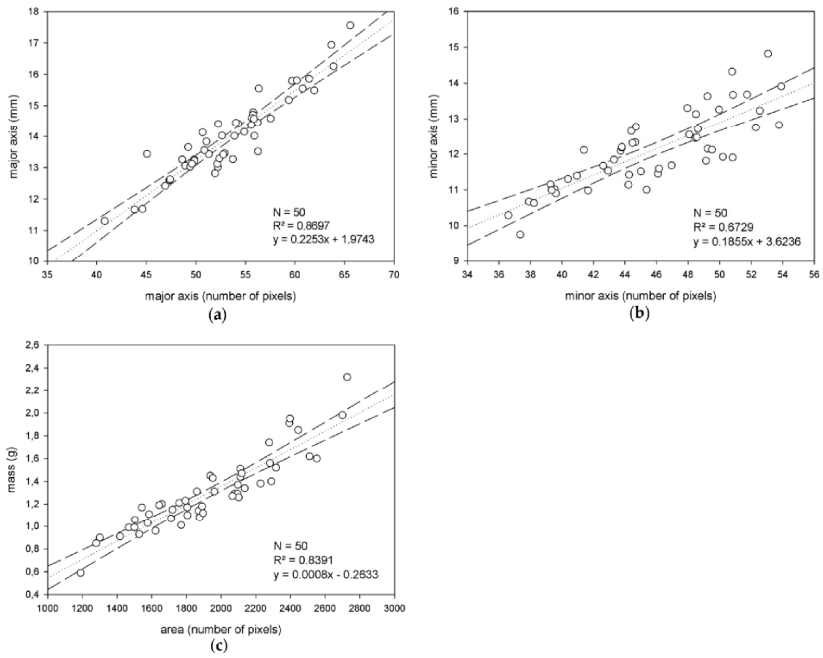


Figure 5. Correlation study performed for the Arbequina variety, considering the three different sizing features of the fruits the experiment is focused on: The major axis (**a**), minor axis (**b**) and mass (**c**).

Globally, there was a positive correlation in all cases, which indicated promising estimation perspectives on the validation sets for all of the features. Nevertheless, there are a few considerations to highlight. First, despite the similar segmentation quality outcomes (Table 1), the correlation results were noticeably better for the Picual variety. Regarding mass modeling, this behavior may be explained by the fact that the method approaches the problem using the projection of the connected components representing the olives onto a 2D plane. Indeed, in that projection, every pixel has the same contribution to the mass of the fruit. This lineal approximation, despite being potentially valid attending to the correlations obtained (to be confirmed later with the validation results), may benefit some varieties more than others depending on their morphological characteristics. Second, to analyze the training correlation results corresponding to the diameters of olives, it is important to underscore that some caution is advised when taking the objective measurements with the digital caliper. Indeed, the lack of firmness of the fruits inevitably induced certain variability in caliper jaws fitting. Since firmness is a defining characteristic of olive fruit varieties, the impact of this variability could be different depending on the kind of fruits under study. This fact has to be considered when assessing the results, and it could partially explain the best correlation for Picual compared to that for Arbequina in terms of size features. On the other hand, the visual determination of the minor axis was found to be more non-specific than for the case of the major axis. Thereby, it also introduced a new variability factor to consider which may explain, at least partially, the correlation differences between the two axes, thus not being univocally imputable to the developments presented here.

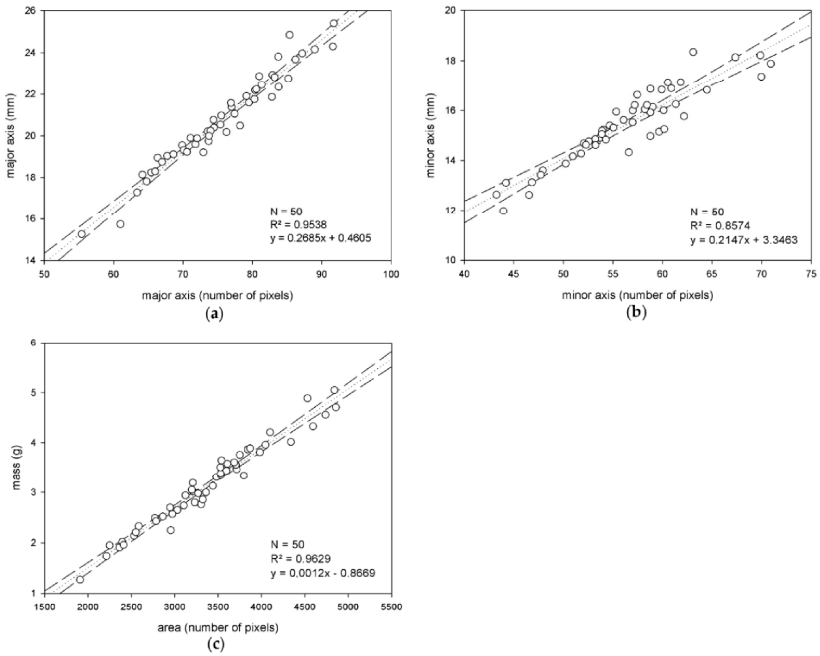


Figure 6. Correlation study performed for the Picual variety, considering the three different sizing features of the fruits the experiment is focused on: The major axis (**a**), minor axis (**b**) and mass (**c**).

The previous approach comprised the development of models specific to each variety. To explore the idea of variety-independent modeling, the two training sets were configured as a unique population to be correlated, thus producing functions applicable to estimate the features of both cultivars. The scatter plots shown in Figure 7 illustrate the training results, which resulted in even higher correlation values than those obtained in the previous case. This could be explained because of a wider domain of the values of the targeted features, which could provide a better adjustment.

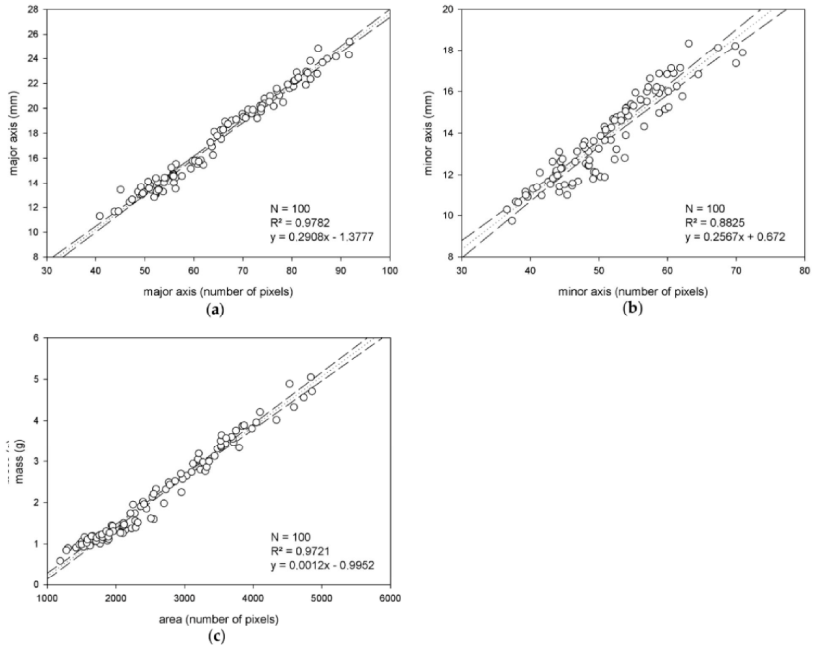


Figure 7. Correlation study of variety-independent model trained on the instances from both Arbequina and Picual varieties, considering the three different targeted features: The major axis (**a**), minor axis (**b**) and mass (**c**).

To evaluate the quality of the different estimation models, the measures proposed before to quantify the differences between the predicted values and the observed ones, have been calculated after applying the models on the corresponding external validation sets. The results can be found below, in Table 2.

Table 2. Estimation results, calculated on the external validation sets, for the three studied features detailed per variety and modeling approach. Outcomes are expressed in terms of Root-Mean-Square Error (RMSE; Equation (12)), Relative Root-Mean-Square Error (SE; Equation (13)), and Relative Mean Error ($|E|$; Equation (14)).

Arbequina Validation Set (N = 150)				
Feature	**Estimation Model**	**RMSE**	**SE (%)**	**|E| (%)**
Major axis	Specific	0.4885 (mm)	3.46	0.86
	Variety-independent	0.5778 (mm)	4.09	0.14
Minor axis	Specific	0.6007 (mm)	4.99	0.09
	Variety-independent	0.7811 (mm)	6.49	2.39
Mass	Specific	0.1220 (g)	9.62	0.78
	Variety-independent	0.1775 (g)	13.99	1.51
Picual Validation Set (N = 150)				
Feature	**Estimation Model**	**RMSE**	**SE (%)**	**|E| (%)**
Major axis	Specific	0.4163 (mm)	1.98	0.03
	Variety-independent	0.4770 (mm)	2.27	0.60
Minor axis	Specific	0.6804 (mm)	4.38	0.29
	Variety-independent	0.8036 (mm)	5.17	1.53
Mass	Specific	0.250 (g)	7.89	2.39
	Variety-independent	0.2439 (g)	7.69	1.65

As can be analyzed, the results produced by the variety-independent models are comparable to those given by the functions specifically trained for each cultivar. This indicates that there is no clear evidence of benefits in the use of specific models to the detriment of the more general solution. This fact increases the expectations on generality and usability of the proposed method. Moreover, the numerical results (with relative mean errors ($|E|$) below 2.5% for all cases; Equation (14)) do not clearly support the need for exploring non-linear solutions, that are far more complex in order to handle and be more sensitive to training populations.

With the aim of determining the probability of getting performance quality for each of the diverse varieties, a one-way analysis of variance was accomplished. To achieve this, the individual relative errors (Equation (15)) produced by the models, for the three features, on the samples in the external validation sets of Arbequina and Picual, were calculated. Then, statistical differences between the populations were studied by mean comparison using the Tukey's test [25] at $p < 0.05$. Table 3 shows the results of this analysis for the two modeling approaches and the three features. In the case of the specific prediction models, no statistical differences were found between Arbequina and Picual for the estimations of the minor axis. Conversely, significant differences were found for the estimations of the major axis and mass. On the other hand, the analysis concluded complementary results for the estimations produced by the variety-independent models. This is, statistically significant differences were found for the minor axis estimations, whilst the major axis and mass estimations verified the null hypothesis. These outcomes reinforce the previous discussion about the suitability of the variety-independent solution.

Table 3. Results of one-way analysis of variance performed on the Arbequina and Picual estimations produced by the specific and variety-independent models. The analyzed populations are the individual relative errors (Equation (15)) produced by the different models on the external validation sets of each variety. The mean ($\overline{X}$) and standard deviation (σ) of each population is given. Dissimilar letters indicate different statistical means according to the analysis of variance using the Tukey's test [25] at $p < 0.05$.

Specific Estimation Models (N = 150)		
Feature	**Arbequina ($\overline{X}$, σ)**	**Picual ($\overline{X}$, σ)**
Major axis	(0.9921, 0.0344) [a]	(1.0005, 0.0199) [a′]
Minor axis	(1.0026, 0.0502) [b]	(1.004, 0.043) [b]
Mass	(0.9985, 0.0946) [c]	(1.0275, 0.0741) [c′]
Variety-Independent Estimation Models (N = 150)		
Feature	**Arbequina ($\overline{X}$, σ)**	**Picual ($\overline{X}$, σ)**
Major axis	(1.0011, 0.042) [a]	(0.9936, 0.0221) [a]
Minor axis	(1.0246, 0.0611) [b]	(0.9847, 0.0486) [b′]
Mass	(1.0068, 0.1428) [c]	(0.9851, 0.0731) [c]

4. Conclusions

In the present paper, a method based on image analysis techniques has been developed for estimating the size and mass of olive fruits. The results underscore the robustness and accuracy of the algorithm this method is based on. Moreover, they support its viability for the development of sorting and grading systems for the olive industry.

In accordance with the results, the segmentation algorithm showed a noticeably good performance in the image segmentation binarization task when compared to ground-truth images. Additionally, it was able to detect the exact number of fruits that appeared in every treated image, thus highlighting the accuracy of the process. It is also interesting to note the steadiness of the method dealing with two different olive fruit varieties, as this increases confidence in its applicability to other cultivars. Nevertheless, future trials will focus on analyzing this aspect of the method to verify this generality. Also, these trials could explore different lightning systems, such as diffuse illumination, which could improve the image acquisition task by minimizing the shadows cast by the fruits, thus enabling more reliable segmentation results.

Regarding the estimation of the major and minor axis, and mass of olives, accurate results were measured, which do not indicate the necessity of exploring non-linear modeling to this effect. Especially remarkable is the analyzed behavior of the variety-independent models, which showed comparable, or even better, performance than specific models. This outcome supports their use in the pursuit of applicability and generalization. Notwithstanding this, future investigations will pursue the verification of this conclusion with studies that include samples from more varieties. Moreover, further and wider investigations will also be conducted to more confidently quantify the impact of pixel weighing linearization for mass estimation.

On the other hand, there is a requirement for the proposed methodology to be applied, in terms of the disposition with the olive fruits that are placed on the images. Thus, it is necessary that a certain minimum distance be maintained between every pair of fruits. This fact does not imply a problem in a real scenario, where a non-flat belt conveyor equipped with cleavages can be used, which provides a way to keep the fruits separated from each other. Nevertheless, further work might explore the enhancement of the image-binarization method presented, with the purpose of making possible a reliable segmentation that will correctly work in a scenario in which olives appear to be touching each other. Notwithstanding, it would probably require a considerable increase in algorithm complexity, so it remains to be determined if it could satisfy the working conditions of a real-time system.

The presented solution comprises a promising starting point to develop sorting and grading technologies based on image analysis, which would provide high value for the olive-manufacturing industry.

Author Contributions: J.M.A. and A.A. conceived, designed and directed the experiments; B.M., A.A. and J.M.P. performed the data and image acquisition; A.A. and J.M.P. designed and developed the methodology. J.M.P. drafted the manuscript, which was revised and edited by J.M.A. and A.A. All authors read and approved the final manuscript.

Funding: The research and APC were funded by the INTERREG Cooperation Program V-A SPAIN-PORTUGAL (POCTEP) 2014–2020, and co-financed with ERDF funds, grant number 0155_TECNOLIVO_6_E, within the scope of the TecnOlivo Project.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Vossen, P. Olive Oil: History, Production, and Characteristics of the World's Classic Oils. *Hortic. Sci.* **2007**, *42*, 1093–1100.
2. Su, C.; Sun, J.; Zhu, W.; Peng, L. History, Distribution, and Potential of the Olive Industry in China: A Review. *Sustainability* **2018**, *10*, 1426. [CrossRef]
3. International Olive Oil Council (IOC). Updates Series of World Statistics on Production, Imports, Exports and Consumption. 2018. Available online: http://www.internationaloliveoil.org/estaticos/view/131-world-olive-oil-figures (accessed on 18 June 2018).
4. United States Department of Agriculture, Foreign Agricultural Service. Oilseeds: World Markets and Trade. Available online: https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf (accessed on 19 June 2018).
5. Russo, C.; Cappelletti, G.M.; Nicoletti, G.M.; Di Noia, A.E.; Michalopoulos, G. Comparison of European olive production systems. *Sustainability* **2016**, *8*, 825. [CrossRef]
6. Connor, D.J.; Gómez-del-Campo, M.; Rousseaux, M.C.; Searles, P.S. Structure, management and productivity of hedgerow olive orchards: A review. *Sci. Hort.* **2014**, *169*, 71–93. [CrossRef]
7. Moreda, G.P.; Ortiz-Cañavate, J.; García-Ramos, F.J.; Ruiz-Altisent, M. Non-destructive technologies for fruit and vegetable size determination—A review. *J. Food Eng.* **2009**, *92*, 119–136. [CrossRef]
8. Sánchez Gómez, A.H.; García García, P.; Rejano Navarro, L. Elaboration of table olives. *Grasas Aceites* **2006**, *57*, 86–94. [CrossRef]
9. Kailis, S.; Harris, D. *Producing Table Olives*; Landlinks Press: Collingwood, Australia, 2007; pp. 173–174, 244–246, ISBN 978-0-643092-03-7.

10. Aquino, A.; Millan, B.; Diago, M.P.; Tardaguila, J. Automated early yield prediction in vineyards from on-the-go image acquisition. *Comput. Electron. Agric.* **2018**, *144*, 26–36. [CrossRef]
11. Mery, D.; Pedreschi, F. Segmentation of colour food images using a robust algorithm. *J. Food Eng.* **2005**, *66*, 353–360. [CrossRef]
12. Cervantes, J.; Taltempa, J.; García-Lamont, F.; Castilla, J.S.R.; Rendon, A.Y.; Jalili, L.D. Comparative Analysis of the Techniques Used in a Recognition System of Plant Leaves. *RIAI—Rev. Iberoam. Autom. Inform. Ind.* **2017**, *14*, 104–114. [CrossRef]
13. Zhang, Y.; Wu, L. Classification of fruits using computer vision and a multiclass support vector machine. *Sensors* **2012**, *12*, 12489–12505. [CrossRef] [PubMed]
14. Sa'ad, F.S.A.; Ibrahim, M.F.; Shakaff, A.Y.; Zakaria, A.; Abdullah, M.Z. Shape and weight grading of mangoes using visible imaging. *Comput. Electron. Agric.* **2015**, *115*, 51–56. [CrossRef]
15. Mizushima, A.; Lu, R. An image segmentation method for apple sorting and grading using support vector machine and Otsu's method. *Comput. Electron. Agric.* **2013**, *94*, 29–37. [CrossRef]
16. Omid, M.; Khojastehnazhand, M.; Tabatabaeefar, A. Estimating volume and mass of citrus fruits by image processing technique. *J. Food Eng.* **2010**, *100*, 315–321. [CrossRef]
17. Diaz, R.; Faus, G.; Blasco, M.; Blasco, J.; Molto, E. The application of a fast algorithm for the classification of olives by machine vision. *Food Res. Int.* **2000**, *33*, 305–309. [CrossRef]
18. Beyaz, A.; Taha, M.; Duygu, İ. Identification of some Spanish olive cultivars using image processing techniques. *Sci. Hortic.* **2017**, *225*, 286–292. [CrossRef]
19. Piras, F.; Grillo, O.; Venora, G.; Lovicu, G.; Campus, M.; Bacchetta, G. Effectiveness of a computer vision technique in the characterization of wild and farmed olives. *Comput. Electron. Agric.* **2016**, *122*, 86–93. [CrossRef]
20. Gatica, G.; Best, S.; Ceroni, J.; Lefranc, G. Olive Fruits Recognition Using Neural Networks. *Procedia Comput. Sci.* **2013**, *17*, 412–419. [CrossRef]
21. Sonka, M.; Hlavac, V.; Boyle, R. *Image Processing, Analysis, and Machine Vision*, 4th ed.; Cengage Learning: Belmont, CA, USA, 2014; pp. 36–38. ISBN 978-1-133-59369-0.
22. Connolly, C.; Fliess, T.A. Study of efficiency and accuracy in the transformation from RGB to CIELAB color space. *IEEE Trans. Image Process.* **1997**, *6*, 1046–1047. [CrossRef] [PubMed]
23. Soille, P. *Morphological Image Analysis—Principles and Applications*, 2nd ed.; Springer: Berlin, Germany, 2013; pp. 63–97. ISBN 978-3-642-07696-1.
24. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cyb.* **1979**, *9*, 62–66. [CrossRef]
25. Tukey, J.W. Comparing individual means in the analysis of variance. *Biometrics* **1949**, 99–114. [CrossRef]

4.1.2. Article 2

Automatic Counting and Individual Size and Mass Estimation of live-Fruits Through Computer Vision Techniques

J. M. Ponce, A. Aquino, B. Millan, J. M. Andújar

Published in:

IEEE Access

Journal: IEEE Access (ISSN: 2169-3536)

Editorial: Institute of Electrical and Electronics Engineers (IEEE)

Reference: vol. 7, pp. 59451-59465

Year: 2019

DOI: 10.1109/ACCESS.2019.2915169

Quality index (Journal Citation Reports®, 2019): 26/90 (Q2) in the category "Telecommunications", 35/156 (Q1) in "Computer Science, Information Systems" and 61/266 (Q1) in "Engineering, Electrical & Electronic". Impact Factor of 3.745.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 7, 2019, accepted May 3, 2019, date of publication May 7, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2915169

Automatic Counting and Individual Size and Mass Estimation of Olive-Fruits Through Computer Vision Techniques

JUAN M. PONCE, ARTURO AQUINO, BORJA MILLAN, AND JOSÉ M. ANDÚJAR, (Senior Member, IEEE)
Department of Electronic Engineering, Computer Systems and Automation, University of Huelva, 21819 Huelva, Spain
Corresponding author: Juan M. Ponce (jmponce.real@diesia.uhu.es)

This work and APC were supported in part by the INTERREG Cooperation Program V-A SPAIN-PORTUGAL (POCTEP) 2014–2020, and in part by the ERDF funds under Grant 0155_TECNOLIVO_6_E, within the scope of the TecnOlivo Project.

ABSTRACT Fruit grading is an essential post-harvest task in the olive industry, where size-and-mass based fruit classification is especially important when processing high-quality table olives. Within this context, this research presents a new methodology aimed at supporting accurate automatic olive-fruit grading by using computer vision techniques and feature modeling. For its development, a total of 3600 olive-fruits from nine varieties were photographed, stochastically distributing the individuals on the scene, using an ad-hoc designed an imaging chamber. Then, an image analysis algorithm, based on mathematical morphology, was designed to individually segment olives and extract descriptive features to estimate their major and minor axes and their mass. Regarding its accuracy for the individual segmentation of olive-fruits, the algorithm was proven through 117 captures containing 11 606 fruits, producing only six fruit-segmentation mistakes. Furthermore, by linearly correlating the data obtained by image analysis and the corresponding reference measurements, models for estimating the three features were computed. Then, the models were tested on 2700 external validation samples, giving relative errors below 0.80% and 1.05% for the estimation of the major and minor axis length for all varieties, respectively. In the case of estimating olive-fruit mass, the models provided relative errors never exceeding 1.16%. The ability of the developed algorithm to individually segment olives stochastically positioned, along with the low error rates of around 1% reported by the estimation models for the three features, makes the methodology a promising alternative to be integrated into a new generation of improved and non-invasive olive classification machines. The new developed system has been registered in the Spanish Patent and Trademark Office with the number P201930297.

INDEX TERMS Computer vision, feature modeling, food industry, fruit grading, image analysis, olive.

I. INTRODUCTION

The olive (Olea europaea) is a species belonging to the family of Oleaceae, which nowadays comprises one of the most significative horticultural crops worldwide. Mainly due to the increase of popularity the olive-derived products have experienced over the last decades, its cultivation, practiced for centuries in the Mediterranean Basin, has spread all around the world. Indeed, it can be found with a growing presence in such disparate countries as China, Australia or USA [1], [2]. This geographical expansion has obviously been accompanied by a huge growth in terms of production, which is directly reflected in the numbers the olive industry deals with. Thus, according to estimations for the 2018/19 crop year, a table olive production of 2,750,000 tons [3], and 3,130,000 tons for the case of olive oil is expected [4].

The increasing demand and consumption of these olive-based products has led the industry to explore the use of new technologies aimed at developing a more profitable, competitive and sustainable market around it. Thus, as in other mainstream crops, olive sector is currently experiencing a process of major transformation [5], [6].

When talking about table olives, fruit classification according to size is a relevant task undertaken during postharvest manufacturing [7]. Size homogeneity, along with other

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

sensory attributes, have a positive impact on the overall consumer opinion about the quality of the product [8]. Furthermore, this is not exclusive of olives, since fruit-size grading is a global issue within food industry when processing high-quality horticultural commodities [9]. So much so that its automation has been historically a challenge to deal with, since this activity has traditionally been performed by hand, with all the drawbacks that it brings. First approaches to the problem have been based on purely mechanical solutions, as the integration of different size hoppers through which fruits can slide, according to their dimensions, into the conveyor belts used during postharvest treatment. However, they can potentially damage fruits [10], [11], since it implies a higher degree of physical manipulation of the commodities. In addition, they present obvious limitations in terms of the features the classification can be based on, and the information that can be recorded after processing.

Within the described scenario, the classical mechanical approaches have been reviewed in recent years [12], being machine vision probably the most investigated technique to build a new generation of less invasive postharvest horticultural classification machines. Thus, Baigvand et al. [13] proposed a machine-vision-based integral solution for dried figs sorting. Sad et al. [14] presented a methodology fusing image processing and supervised machine learning for grading mangoes according to shape and mass. Focused on the same crop, Wang et al. [15] suggested the use of RGB-D sensing for in-field fruit size estimation. In addition, Mizushima and Lu [16] faced the segmentation of images of apples, in order to enable their automatic sorting.

Likewise, the use of computer vision techniques has already been approached within the olive sector. Gatica et al. [17] proposed RGB-image analysis and the use of neural networks to recognize the fruits directly in the trees, thus ideally estimating the best harvest time. On the other hand, several studies have focused on varietal identification on the basis of image analysis. Thus, Martínez et al. [18] approached the problem by feature extraction from images of olive endocarps, and then using partial least square-discriminant classifiers. Similarly, the proposal by Beyaz et al. [19] used captures of fruits and endocarps to identify olive cultivars. Aside from variety classification, research has also been conducted to deal with defective fruit discrimination. To this effect, Diaz et al. [20] compared different algorithms for categorizing olives according to their surface condition. On the other hand, Puerto et al. [21] presented a methodology for differentiating olives collected from the ground from those harvested directly from the trees, as the former impoverishes quality of the subsequently produced olive oil.

In a previous research, Ponce et al. [22] successfully found strong evidences about the viability of the estimation of mass and size of olives by means of image analysis. In order to extend the scope approached then, a highly improved new methodology, based on computer vision and feature modelling, is proposed in the present paper. It undertakes a new scenario, closer to match the needs of the olive industry in terms of automated fruit grading. Thus, an image acquisition chamber, potentially integrable in a conveyor belt, was designed with the capability of taking photos of olive batches under controlled conditions. With this device, groups of olives from nine different varieties were photographed, stochastically positioning the individuals on the capturing area to mimic the chaotic distribution they would have on real a conveyor belt. Then, the images were binarized by clustering-based image thresholding, and the olives appearing fused for being in touch were separated employing mathematical morphology principles. After isolating the individual olives, descriptive data of the features under study were extracted from the transformed images. Finally, a linear regression analysis of the correlation between these measurements and their corresponding individual objective observations, previously taken in laboratory, was performed to compute the models for estimating size and mass of every single olive appearing in images.

Unlike other studies in this field, the methodology presented is novel because, in the first place, it proposes a specific image acquisition system, designed to favor binarization of the captures with high precision, at a low computational cost. It makes possible to face the subsequent individual segmentation of the fruits within a timeframe potentially assumable to be applicable in a real time system, with the implications that this entails in terms of usability for the olive industry. In addition, the significant volume of samples used during the development of the study offers guarantees of the reliability and generality of the results obtained.

Hereafter, the manuscript is divided into three main parts. First, section II focuses on the experimental design, presenting different aspects related to the characteristics of those fruit- samples used throughout the research (section II-A), describing how the image acquisition and reference data collection (section II-B) was carried out, and detailing the framework whereby the image analysis algorithm was developed (section II-C). Section III offers insights into the developed methodology, laying emphasis on how this image analysis procedure was designed to binarize (section III-A), segment at fruit-level (section III-B), and postprocess the initial captures (section III-C). Next, feature characterization (section III-D) and modelling (III-E) are addressed, and methodology's performance evaluation is detailed (section III-F). In section IV, the results attained are presented and discussed. Finally, the last section summarizes the findings achieved, and formulates the main conclusions derived from the present study.

II. MATERIALS AND METHODS

A. SAMPLE COLLECTION

Nine Olea europaea varieties were considered for this study: Arbequina, Arbosana, Picual, Ocal, Changlot Real, Verdial de Huévar, Lechín de Sevilla and two experimental ones, named as 967 and 1030, respectively. Thus, 400 samples per variety (3,600 in total) were gathered by hand in olive

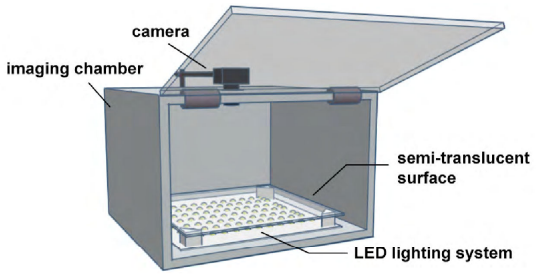


FIGURE 1. Image acquisition chamber.

orchards located in Gibraleón (37°20′09.2″N 7°02′19.8″W), province of Huelva (Andalusia, Spain), in October 2018. For each variety, samples were selected to cover the observed variability in terms of size.

B. REFERENCE DATA AND IMAGE ACQUISITION

For image capture, an imaging chamber was designed and crafted with the aim of isolating the scene from any external light source, thus maximizing illumination control (see Fig. 1).

The fruits were placed on a semi translucent white plastic sheet. This sheet, with dimensions of 500 × 500 × 2 mm, was neatly disposed 65 mm from the bottom of the chamber, and it was illuminated from below by a set of seven equally distributed strips composed of 25 LEDs of 5V each. Due to this lighting system design, it was possible to avoid shadows cast by the fruits. As capturing device, a Sony α7-II digital mirrorless camera (Sony Corp., Tokyo, Japan) was installed at the top of the chamber, looking perpendicularly at the imaging area. The camera mounted a 24 Mpx CCD stabilized sensor, and it was equipped with a Zeiss 24/70mm lens with optical stabilization. It was set in manual mode, configuring the aperture in f/7.1, shutter speed in 1/50s, ISO sensitivity in 250, and focal length in 31mm. The camera was set to save images in JPEG file format, with a 6000 × 3376-pixel resolution, a color-depth of 24 bits, and a pixel density of 350 ppi.

Every variety set of 400 olives was divided into batches of 50 fruits, which were then photographed separately, thus obtaining eight images per variety (72 in total). The only criterion regarding to the way the fruits where disposed to be imaged, was to force multiple touching occurrences. This prerequisite was established with the purpose of approaching a complex and realistic scenario.

Additionally, in order to test the robustness and accuracy of the developed image analysis algorithm, a set of five stress-images were acquired per variety (45 in total). The number of fruits to appear was progressively incremented in each of these five images. Two examples of the images acquired can be checked in Figure 2. Furthermore, Table 1 summarizes how the olive-fruit samples were organized and photographed for the nine studied varieties.

Once a batch was photographed, measurements of mass and size were conducted and registered for every individual

FIGURE 2. Examples of captured images for Picual variety: (a) regular image; (b) stress image.

TABLE 1. Materials: organization of olive-fruit samples and images for each of the nine varieties used.

	Olive fruits		Images acquired	
Variety sample	Training subset	External validation subset	Regular images	Stress images
400 (fruits)	100 (fruits)	300 (fruits)	8[a]	5[b]

[a] 50 fruits per image.
[b] Stochastic number of fruits.

fruit. To do so, a KERN PCB 3500-2 precision balance (KERN & Sohn GmbH, Balingen, Germany) was used to assess olive mass (in grams - g). To grade the size of each fruit, its corresponding major and minor axis length were measured (in millimeters - mm) using a Digital Vernier Caliper, which provided 0.01mm of resolution and 0.02mm of accuracy. It must be pointed out that, for every variety, the major and minor axes of 100 of the 400 individuals (900 in total), were independently measured by three different observers, again manually by using a digital caliper. This, in order to later perform a study of variability, with the goal of assessing the degree of uncertainty introduced by measuring the size of the olives by hand, using the digital caliper.

C. IMAGE ANALYSIS ALGORITHM IMPLEMENTATION

The core of the proposed methodology is an image analysis algorithm designed in order to, first, transform the olive-fruit captures into binary-segmented images, and second, to extract from them data structures with which to perform fruit counting and characterization of the three different features to estimate. The latter, for each of the fruits that appear in the initial captures. Its development was carried out by mainly using transformations based on mathematical morphology, and binarization by statistical thresholding techniques. Regarding its implementation, MATLAB and the Image Processing Toolbox Release 2018a (The MathWorks, Inc., Natick, Massachusetts, USA) were used. To note this technology was chosen only for prototyping purposes, with the awareness that a real system, based on the proposed methodology, would require an algorithmic implementation based on a computationally more flexible and efficient language.

III. DEVELOPED METHODOLOGY

The diagram shown in Fig. 3 illustrates the methodology resulting from the conducted research. Essentially, the dataset

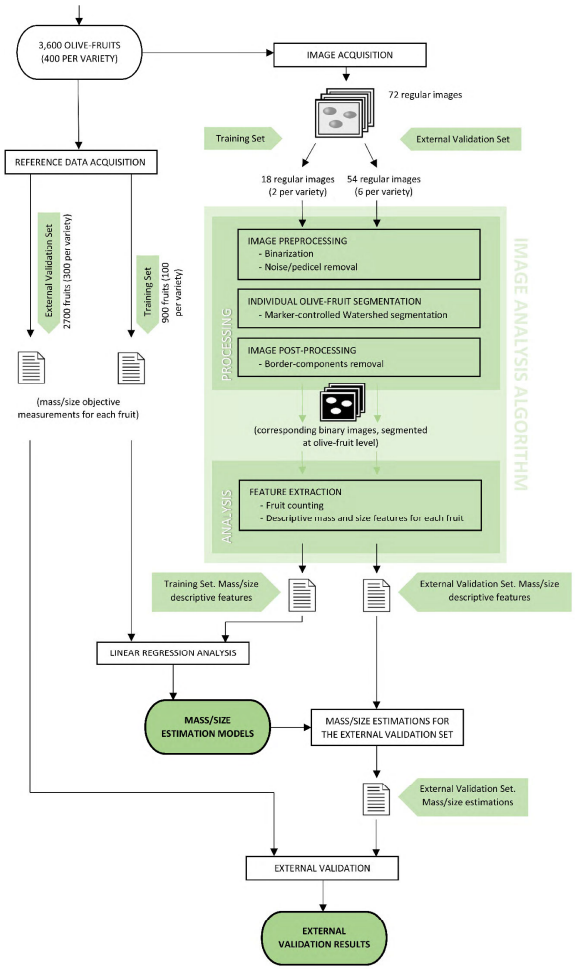


FIGURE 3. Representative diagram of the developed methodology for computing and validating the mass and size olive-fruits estimation models.

composed of the acquired olive-fruit images and their corresponding reference measurements, previously recorded, is used throughout an image processing and analysis procedure aimed at computing, and ultimately validating, the pursued models for fruit mass and size estimation. This process of transforming the captures and extracting descriptive features of fruits, to feed estimation model training, is described in detail throughout the subsections below.

A. IMAGE PREPROCESSING

As it has been stated before, the present study is focused on a scenario in which olives may appear in touch with other in the images. Within this context, the developed image analysis algorithm was designed on the basis of the Watershed transform [23] for individually segmenting olive-fruits. This section describes a set of image transformations applied to favor the performance of this subsequent Watershed transform application.

First, an image binarization consisting in isolating those pixels corresponding to fruits from the background is carried out. To that end, images are previously transformed from RGB to HSV color space [24]. A large body of literature has investigated the importance of using the right color representation when developing image analysis procedures, with no exception when treating captures of horticultural products, whatever the pursued objectives are [25], [26]. In the present study, the specific illumination conditions enabled the acquisition of images with an important differentiation, in terms of luminosity or brightness, between the background-pixels and the fruits. In order to exploit this feature to the purpose of binarizing the image, HSV comprises a suitable color representation, since brightness information can be directly analyzed in the specific channel V (Value). Thus, for its processing, this channel is treated as a grayscale image, denoted as I_V, and its grey-level values are inverted to represent background pixels with lower values than those of fruits:

$$\bar{I_V} = 255 - I_V \tag{1}$$

where 255 is the highest possible grey-level value for images with 8 bits-per-channel of color depth. Then, $\bar{I_V}$ is binarized based on a threshold computed by Otsu's method [27]. This global thresholding technique assumes that there are two classes of pixels in terms of their grey-level values, those belonging to the foreground and those which correspond to the background. Thus, the method automatically calculates the optimum threshold for class separation by maximizing the inter-class variance or, analogously, by minimizing the intra-class variance, so the sum of the measured spreads of the pixel levels of both classes is minimum. Hence, the threshold Th_{Otsu} is computed for image $\bar{I_V}$, which is then binarized as:

$$I_{Bin}(x, y) = \begin{cases} 255 & if\ \bar{I_V}\,(x, y) > Th_{Otsu} \\ 0 & in\ any\ other\ case \end{cases} \tag{2}$$

Next, a morphological opening is applied to the resulting binary image with a double purpose: 1) to eliminate tiny groups of white pixels which cannot represent olives because of their anomalous dimensions; 2) to remove the olive pedicels that might be present in the image. Mathematically:

$$I_{Bin2} = \gamma_\beta\,(I_{Bin})\,, \tag{3}$$

where γ is the morphological opening that uses the disk-shaped structuring element β of 30 pixels in radio [23, pp. 106-108]. This process of segmentation and noise/pedicel removal is illustrated in Fig. 4.

As it can be checked, I_{Bin2} is an accurate segmentation of olives from the background. However, olives in touch appear fused building enlarged connected components (sets of neighbor pixels). The aim of the procedure described in section III-B will be to separate olives fused in I_{Bin2}, so as to every connected component corresponds to a single olive. With this goal, preprocessing finishes by computing the

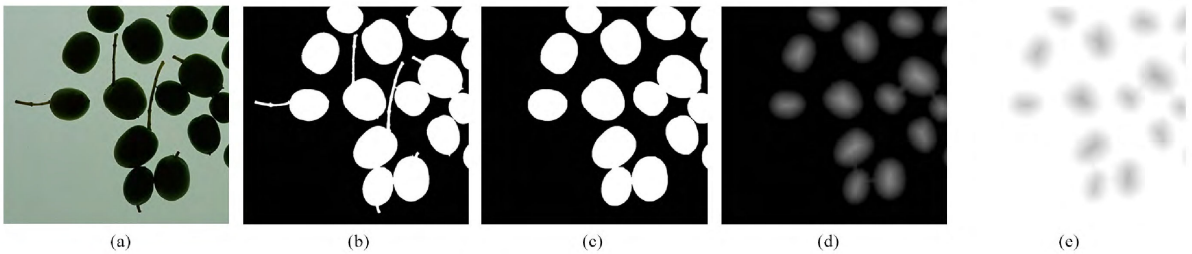


FIGURE 4. Illustration of the image preprocessing: (a) original sub-image; (b) binary sub-image; (c) binary sub-image after noise and pedicel removal; (d) distance transform of image (c); (e) complement of the image (d).

distance transform DT on I_{Bin2}, which calculates, for every pixel, its Euclidean distance to its nearest background pixel:

$$I_{DT} = DT\left(I_{Bin2}\right), \tag{4}$$

where, DT is mathematically formulated as:

$$[DT(I_{Bin2})](x,y) = min\left\{\sqrt{(x-x')^2+(y-y')^2} \mid I_{Bin2}\left(x',y'\right)=0\right\} \tag{5}$$

This information will be key for olive separation in I_{Bin2}. Indeed, I_{DT} is a meaningful and simplified image in which the mass center of all olives, originally in touch or not, constitute a regional maximum for being the local farthest point to the background (see Fig. 4-(d)). Note that a generic regional maximum M of an image can be defined as a set of neighbor pixels with a given grey-level value v, such that every pixel in the neighborhood of M has a value strictly lower than v. The entire preprocessing is illustrated in Fig. 4.

B. INDIVIDUAL OLIVE SEGMENTATION

The goal at this stage is the accurate application of the Watershed transform to individually segment olives in image I_{Bin2}. Conceptually, the Watershed approximates an image as a topographic surface, where grey-level values represent altitudes, being the darker regions taken as catchment basins. Then, the surface is flooded, and the basins are divided by water convergence lines, whose path is influenced by the deepness, size and shape of basins. Formally, the definition and implementation of the Watershed transform is not obvious and has produced a great amount of specific literature over the years, which is encouraged to be consulted for deeper study [28], [29].

According to the given explanation, if the complement image of I_{DT}, $\overline{I_{DT}} = 255 - I_{DT}$ (see Fig. 4-(e)), is seen as a topographic surface, olives constitute catchment basins. Additionally, these basins are ideally divided by draining lines traced according to the influence produced by the characteristics of every basin with respect to its surroundings. However, the direct application of the Watershed to $\overline{I_{DT}}$ would produce oversegmentation with high probability, as it is well-known to be very sensitive to local irregularities [29]. To overcome this difficulty, a marker-controlled Watershed segmentation is proposed [30]. A marker is a connected component belonging to the image to be segmented. Thus, internal markers, those inside olives, and external markers, belonging to the background, are found. Then, they are used to limit the regions allowable by the watershed to those of olives, and to accordingly modify the gradient of the image.

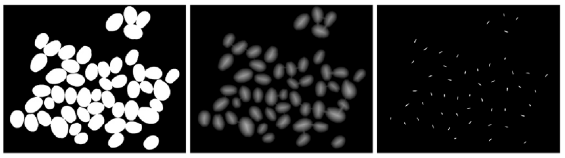


FIGURE 5. (a) starting binary image, I_{Bin2}; (b) image resulting after applying the distance transform on (a), I_{DT}; (c) extended regional maxima of (b), $I_{RMaxBin}$.

As a first step to obtain the set of internal markers, the regional maxima located at the olive centres in image I_{DT} are extracted by applying the *h-maxima* transform [23, pp. 201-204]. This transform firstly removes irrelevant regional maxima from the image, which will be those with a height lower or equal than h. Mathematically:

$$I_{HMax} = R^{\delta}_{I_{DT}}\left(I_{DT} - h\right), \tag{6}$$

where, generically, $R^{\delta}_{I_{Mask}}(I_{Marker})$ is the morphological reconstruction by dilation of I_{Mask} from marker I_{Marker}, using a unitary structuring element [31]. Respecting parameter h, a value of 5 was enough to retain only significative maxima (the choice of this parameter value was proven to have a wide range of optimum values in this particular case). Next, the surviving relevant regional maxima in image I_{HMax} are extracted by computing:

$$I_{RMax} = I_{HMax} - R^{\delta}_{I_{HMax}}(I_{HMax} - 1), \tag{7}$$

and thresholding the resulting image:

$$I_{RMaxBin}(x,y) = \begin{cases} 255 & if\ I_{RMax}\,(x,y) > 0 \\ 0 & in\ any\ other\ case, \end{cases} \tag{8}$$

The result of operations (6)-(8) can be consulted in Fig. 5.

At this point, the exact number N of fruits present in the image, originally in touch or not, is obtained by counting

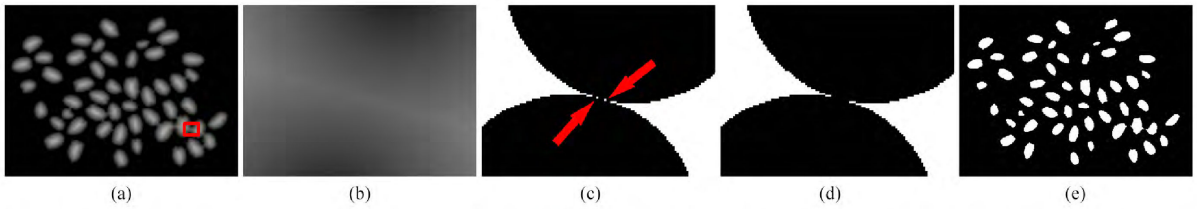


FIGURE 6. Generation of the image of internal markers, I_{IntMk}: (a) image of the distance function; (b) close-up of the red-squared zone in image (a); (c) binarization of (b) with a threshold Th_{Min} value of 65; (d) removal of the artificial connected components thanks to operation (9); (e) Image I_{IntMk}, where each connected component represents an internal marker for the application of the watershed transform.

the number of connected components present in $I_{RMaxBin}$. Next, this information is exploited to finally decide the set of internal markers, by building a binary image in which all olive fruits are disconnected, but in such a way that they keep their spatial relationship/influence; note that this constraint preserves the original major draining lines. To do so, the objective is finding the minimum threshold value, Th_{Min}, producing a binarization of the distance transform image, I_{DT}, exactly containing N connected components corresponding to the olives disaggregated. Indeed, assuming that the image contains olives in touch, the lower the threshold value is, the lower the number of connected components contains the binary image as they tend to fuse building aggregations. Contrary, as the threshold value increases, the olives tend to disconnect, and the number of connected components converge to N. However, there are two situations to consider in the solution for finding Th_{Min}: 1) smaller olives may disappear when binarizing image I_{DT} from a certain threshold value, depending on the size of the larger ones and the size of aggregations; 2) tiny artificial connected components may appear when binarizing I_{DT} as a consequence of the discrete and quantized nature of the image (check Fig. 6-(a) to (c)).

Therefore, the mathematical solution to the described problem, which provides the set of internal markers, can be formulated as follows:

$$I_{IntMk} = R^{\delta}_{I_{Cand}}\left(I_{RMaxBin}\right), \tag{9}$$

where I_{IntMk} contains N connected components corresponding to the individual olives, being

$$I_{Cand} = max\{I_{ThMin}, I_{RMaxBin}\} \tag{10}$$

Indeed, the morphological reconstruction planned in (9) discards the previously pointed out artificial connected components probably present in I_{Cand} (check Fig. 6-(c)-(e)). Additionally, (10) recovers in I_{Cand} the smaller connected components that may have disappeared when calculating I_{ThMin}. Thus, I_{ThMin} is the binarization of the distance transform image I_{DT}:

$$I_{ThMin}(x,y) = \begin{cases} 255 & if\ I_{DT}(x,y) > ThMin \\ 0 & in\ any\ other\ case \end{cases} \tag{11}$$

when $ThMin$ takes the minimum threshold value for which I_{IntMk} contains N connected components (given by the generic function CC),

$$Th_{Min} = min\{t \mid CC\left(I_{IntMk}\right) = N, t = 0, \ldots, 254\} \tag{12}$$

Consequently, in order to computationally determine Th_{Min}, an incremental iterative binarization over every possible threshold value starting from 1 is performed. Thus, by binarizing the distance-transform image I_{DT} at every value, Th_{Min} is decided when the resulting image contains an equal number of connected components, N, than the regional maxima image, I_{IntMk}.

Next image $I_{DT_{IntMk}}$ resulting from the application of the distance transform (5) to I_{IntMk} is calculated (Fig 7-(a)). Then, the set of external markers is obtained by applying the Watershed transform to the complement of $I_{DT_{IntMk}}$, $\overline{I_{DT_{IntMk}}}$, as follows:

$$I_{ExtMk}(x,y) = \begin{cases} 0 & if\ [WS(\overline{I_{DT_{IntMk}}})](x,y) = 0 \\ 1 & in\ any\ other\ case \end{cases}, \tag{13}$$

where WS stands for Watershed transform. The steps to obtain this marker-image are visually represented in Fig. 7. Note that the external markers are the dark lines illustrated in Fig. 7–(c).

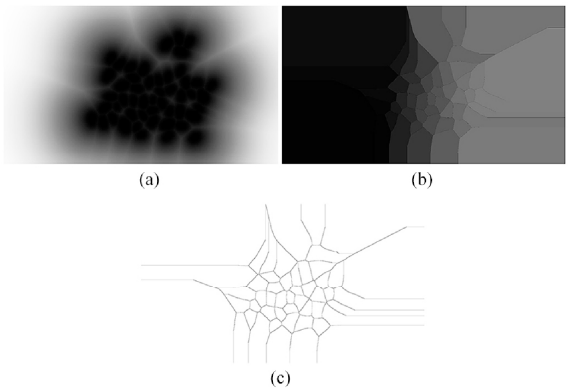


FIGURE 7. Computation of the set of external markers $I_{ExtMark}$: (a) $I_{DT_{IntMk}}$, which results from inverting the image resulting from the application of the distance transform to the image of internal markers $I_{IntMark}$; (b) image $WS(\overline{I_{DT_{IntMk}}})$, resulting from the application of the watershed transform to $\overline{I_{DT_{IntMk}}}$ (a); (c) image of external markers (dark lines), $I_{ExtMark}$, resulting from binarizing $WS(\overline{I_{DT_{IntMk}}})$ (b).

At this point, it should be noted that the set of lines integrating the external markers offer a partial solution to the separation problem, but it often provokes oversegmentation. Thus, the external markers are used to force the occurrence of regional minima (operation known as minima imposition [23, pp. 172-173]) in the complement distance transform image $\overline{I_{DT}}$, as:

$$I_{MinImp} = R^{\varepsilon}_{min\{\overline{I_{DT}}, I_{ExtMark}\}}(I_{ExtMark}) \tag{14}$$

where, generically, $R^{\varepsilon}_{I_{Mask}}(I_{Marker})$ is the morphological reconstruction by erosion of I_{Mask} from marker I_{Marker}, using a unitary structuring element [31]. After this minima imposition, the Watershed transform is applied to I_{MinImp}, thus providing the definitive set of ridge lines used to eventually separate the olives in I_{Bin2} as:

$$I_{Final} = min\{I_{Bin2}, I_{Lines}\}, \tag{15}$$

where

$$I_{Lines}(x, y) = \begin{cases} 0 & if\ [WS\ (\overline{I_{MinImp}})]\ (x, y)] = 0 \\ 1 & in\ any\ other\ case \end{cases}, \tag{16}$$

The definitive outcome at this stage, I_{Final}, is shown in Fig. 8.

C. IMAGE POSTPROCESSING

Once object separation is finished, tiny artificial connected components may have been generated as a consequence of the application of the Watershed transform. These exceptional oversegmentation occurrences are removed by computing a morphological opening, followed by a reconstruction of the resulting image to restore the exact size and shape of the surviving connected components. Mathematically:

$$I_{Olives} = R^{\delta}_{I_{Final}}(I_{Aux}) \tag{17}$$

$$I_{Aux} = \gamma_{\beta}(I_{Final}) \tag{18}$$

β represents the disk-shaped structuring element used in the opening operation. Its radius length is adaptatively computed per variety as follows:

$$\beta = \overline{r^v} - 5 \times \sigma_r^v - 1, \tag{19}$$

where $\overline{r^v}$ and σ_r^v are, respectively, the mean and the sample standard deviation values defining the statistical distribution of observations of the minor radius length of the connected components present in I_{Final}. The distributions were studied for each variety v by analyzing all the images with the methodology described in sections III-A and III-B, and using the procedure detailed in section III-D to calculate the radius length values. Hence, the distribution of olive sizes per variety is assumed to be Gaussian, and only extremely non-representative examples, those out of the average minus five times the standard deviation, are discarded for being too tiny to be meaningful.

Finally, postprocessing ends with removing from I_{Olives} those olives not completely contained in the image. To do so, those connected components reaching the border in I_{Olives} are discarded by performing:

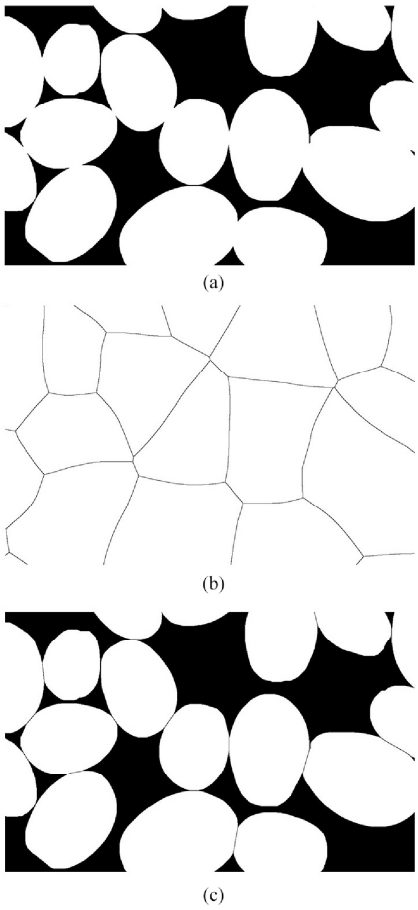


FIGURE 8. Illustration of final olive separation: (a) binary image I_{Bin2} with olive aggregations to be separated; (b) ridge lines obtained by marked-controlled watershed segmentation; (c) binary image, after olive-fruit separation, which is calculated as the minimum value of (a) and (b) for every pixel.

$$I_{OlivesDef} = I_{Olives} - R^{\delta}_{I_{Olives}}(I_{BorderSeeds}), \tag{20}$$

where $I_{BorderSeeds}$ contains seeds for each olive connected to the image border

$$I_{BorderSeeds} = min\{I_{Olives}, I_{Border}\}, \tag{21}$$

and I_{Border} is a border image

$$I_{Border}(x, y) = \begin{cases} 255 & if\ \ (x, y)\ is\ a\ border\ pixels \\ 0 & in\ any\ other\ case \end{cases}\Big\} \tag{22}$$

This subprocess is illustrated in Fig. 9 by treating a sub-image instead of a full capture. This, in order to make easier its visualization.

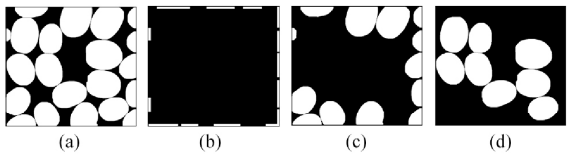


FIGURE 9. Illustration of the removal of the olives that are not completely contained in the image: (a) binary image after individual olive segmentation; (b) seeds corresponding to each of the olives connected to the border; (c) components corresponding to the olives connected to the border; (d) result after components in (c) are removed from (a).

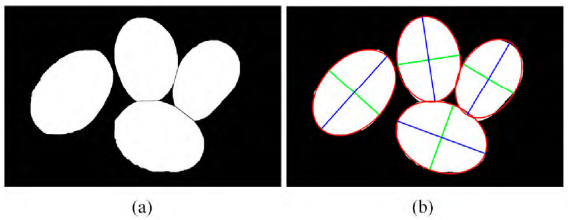


FIGURE 10. Representation of the ellipses, and their corresponding axes, computed for characterizing olive-fruit size.

It should be noted that in Fig. 9-(b), seeds have been expanded and exaggerated for the purpose of facilitating its visual representation.

D. OLIVE SIZE AND MASS CHARACTERISATION

The image analysis methodology presented individually segmented olive-fruits following the principles specified in section III-B. At this point, the objective is the formulation of descriptive features, measurable in the segmented images, to approach the major and minor axis length, and mass, of every individual olive. Finally, these features are exploited to build estimation models.

First, the major and minor axis length of an olive are proposed to be characterized by the length in pixels of the two axes of the ellipse having the same normalized second central moment [32] than its connected component:

$$MajAx\left(CC_i\right) = \max\left(E_{D_1}\left(CC_i\right), E_{D_2}\left(CC_i\right)\right), \quad CC_i \subseteq I_{Olives}, \tag{23}$$

$$MinAx\left(CC_i\right) = \min\left(E_{D_1}(CC_i), E_{D_2}(CC_i)\right), \quad CC_i \subseteq I_{Olives}, \tag{24}$$

E_{D_1} and E_{D_2} denotes the diameters, D_1 and D_2, of the two axes of ellipse E fulfilling the previous definition for connected component CC_i of image I_{Olives}. This approach is illustrated in Fig. 10.

Second, for characterizing mass of an olive, the area in pixels of its corresponding connected component is considered. The first approach to this proposal is to assume the extent to which each pixel equally contributes to the mass characterization, independently of its position within the connected component. Mathematically:

$$Area\left(CC_i\right) = \#CC_i, \quad CC_i \subseteq I_{Olives}, \tag{25}$$

where # stands for the cardinal operation.

Additionally, a set of alternative approaches, based on pixel weighing, are proposed as well. It should be noted that a connected component is, in fact, a 2D projection of an olive, which actually has an ellipsoidal three-dimensional body. In order to introduce this notion of three-dimensionality, it seems reasonable to consider that the contribution of a given pixel, when approximating the mass of the fruit, must be greater the closer this pixel is to the mass center of the corresponding connected component. This argument suggests weighing the contribution of every pixel according to its relative location within the connected component which it belongs to.

Therefore, it is proposed a pixel-weighing scheme by means of the application of the distance transform in conjunction with a set of functions, aimed to express the pursued three-dimensionality.

Hence, the area of a connected component CC_i contained in image I_{Olives}, is calculated using a generic weighing function w, as:

$$Area_w(CC_i) = \sum\nolimits_{(x,y)\in CC_i} w\left(\left[NDT_v\left(I_{Olives}\right)\right](x,y)\right), \quad CC_i \subseteq I_{Olives} \in v \tag{26}$$

NDT_v refers to the distance transform function (4), normalized by the maximum olive-to-background pixel distance measured throughout the whole set of processed images, for the corresponding olive variety v (27), as shown at the bottom of this page.

The purpose of the weighing function is to modulate the initial weights, provided by the application of the distance transform, according to the ellipsoidal shape of the fruits. To this end, logarithm- and root-based functions are selected and studied, since the draw similar shapes to that observable in a cross-sectional view of an olive.

Therefore, the proposed weighing functions w, and the labels they will be referred with from now on, are defined as:

- Natural logarithm (NLog):

$$w\left(p\right) = ln(1 + (e-1) \times p) \tag{28}$$

- Logarithm to base n (Log_n):

$$w\left(p\right) = \log_n\left(1 + (n-1) \times p\right), \quad n = 2, 3, 5, 10 \tag{29}$$

- Square root (Root_n):

$$w\left(p\right) = \sqrt[n]{p}, \quad n = 2, 3, 5, 7, 10 \tag{30}$$

The representation as an image of the application of the different weighing functions is illustrated in Fig. 11, where the weigh assigned to each pixel is represented as a grey-level value. All these alternatives will later be evaluated and compared.

E. OLIVE SIZE AND MASS MODEL TRAINING

For training estimation models for olive minor and major axis length, and mass, a set of 100 olive-fruits was built per variety

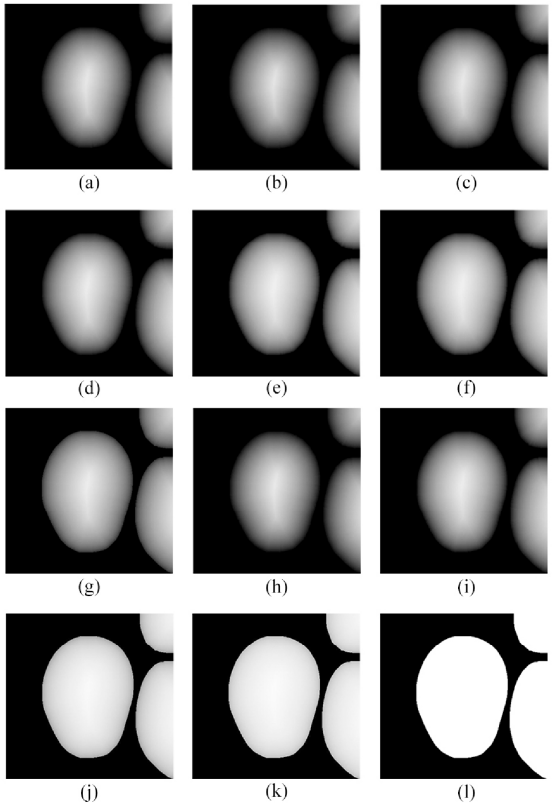


FIGURE 11. Pixel-weighing representation for each of the mass models considered: (a) NLog; (b) Log_2; (c) Log_3; (d) Log_5; (e) Log_7; (f) Log_{10}; (g) $Root_2$; (h) $Root_3$; (i) $Root_5$; (j) $Root_7$; (k) $Root_{10}$; (l) representation of equally-weighed pixels.

(900 in total), thus keeping disjoint sets of 300 individuals per variety for external validation (2,700 in total). The sets for training were made up with samples representing the particular observed variability for each olive variety in terms of size and mass. On the other hand, the images containing the training olive samples were analyzed, and the features described above were calculated for all of them. Then, a linear regression analysis was performed per variety on the training instances. It allowed to model the relationship between the magnitude of the features given by image analysis and the corresponding objective data previously registered in laboratory. This training scheme produced three estimation models per variety, one per feature, which will be referred as variety-dependent estimation models hereafter. In addition, the linear regression analysis was replicated by considering the nine different training sets as a whole. It made possible to generate and assess the behavior of a unique set of three estimation models applicable to all varieties. These models will be called variety-independent estimation models from here on.

F. APPROACH FOR THE METHODOLOGY'S PERFORMANCE EVALUATION

The performance of the proposed methodology was assessed following a double approach.

First, the ability of the image analysis algorithm to individually segment olives was studied. For this purpose, the 72 regular images (8 per variety), containing 50 olives each, were processed with the algorithm, and the number of individual olives found was evaluated. Additionally, the algorithm was also challenged under this criterion by analyzing the set of 45 stress images (5 per variety).

Second, the behavior of the whole methodology, comprising the image analysis algorithm and the different estimation models, was assessed. For each of the three modelled features, olive major and minor axis length, and mass, estimations given by the methodology were confronted to the reference values obtained in laboratory using the following proposed metrics:

- Root-Mean-Square Error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{31}$$

- Relative Root-Mean-Square Error (expressed as percentage):

$$SE = \frac{RMSE}{\frac{\sum_{i=1}^{n} y_i}{n}} \times 100 \tag{32}$$

- Relative Mean Error (expressed as percentage):

$$|E| = \frac{\left|\sum_{i=1}^{n} (\hat{y}_i - y_i)\right|}{\sum_{i=1}^{n} y_i} \times 100 \tag{33}$$

where, for the *i-th* fruit of a set of n elements, $\hat{y}_i$ refers to the predicted value and y_i stands for its corresponding reference measure. Additionally, one-way analysis of variance was also conducted on the estimation results of the different developed models for the nine varieties. Mean comparison was attempted, performing the Tukey's test [33] at $p < 0.05$, on the population of individual relative errors, e_i, defined as the ratio between the estimated and the actual value considered:

$$e_i = \frac{\hat{y}_i}{y_i} \tag{34}$$

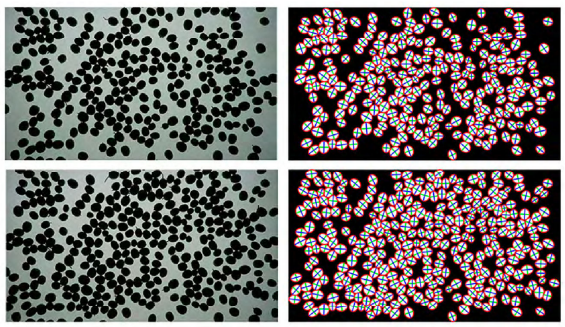
At this point, it is important to recall that the minor and major axis length of 100 olives per variety (900 in total) were measured by three different observers. For these samples and

$$[NDT_v(I_{Olives})](x,y) = \frac{[DT\,(I_{Olives})]\,(x,y)}{max\left\{\sqrt{(x'-x'')^2 + (y'-y'')^2} \mid I'_{Olives}(x',y') = 255 \wedge I''_{Olives}(x'',y'') = 0\right\}}, \tag{27}$$

TABLE 2. Fruit-detection accuracy tested on stress images.

Variety	Stress Images					
	stress_image_0 (detected/actual fruits)	stress_image_1 (detected/actual fruits)	stress_image_2 (detected/actual fruits)	stress_image_3 (detected/actual fruits)	stress_image_4 (detected/actual fruits)	Overall (detected/actual fruits success rate)
967	103/103	165/165	187/187	254/254	311/311	1020/1020
1030	138/138	182/182	210/210	246/246	323/323	1099/1099
Arbequina	158/158	194/194	221/221	251/249	304/303	1128/1125
Arbosana	111/111	137/137	190/190	250/250	355/355	1043/1043
Changlot	66/66	91/91	117/117	156/156	237/237	667/667
Lechín	71/71	112/112	153/153	211/211	258/257[b]	805/804
Ocal	95/95	100/100	113/113	117/117	141/141	566/566
Picual	164/164	195/195	207/207[a, b]	233/233	256/256	1055/1055
Verdial	64/64	91/91	112/112	140/140	220/220	627/627
	-	-	-	-	-	8110/8106

[a] Denotes underestimation occurrence.
[b] Denotes overestimation occurrence.



(a) (b)

FIGURE 12. Examples of stress-images captured for variety 967 (a), and the corresponding results after processing/analysis algorithm is applied (b).

features, the average of the three independent observations was used as reference measure y_i for the application of the metrics above. Notwithstanding, the motivation for performing this multiple evaluation for a subset of olives, was to investigate the degree of uncertainty induced by the observer when measuring the olive axis with a digital caliper. With this purpose, the *SME* metric was defined as:

$$SME = \frac{\sum_{i=1}^{n} ISME_i}{n} \times 100, \tag{35}$$

where

$$ISME_i = \frac{max\{|y_{i,j} - \overline{y_i}|\}}{\overline{y_i}} \tag{36}$$

Indeed, for every fruit i involved in the study, an error $ISME_i$ was defined as the absolute maximum deviation of the three observations $y_{i,j}, j = 1, \ldots, 3$, with respect to the mean value of the observations $\overline{y_i}$.

IV. RESULTS AND DISCUSSIONS

A. RESULTS OF THE IMAGE ANALYSIS ALGORITHM

The image analysis algorithm showed 100% of accuracy when individually segmenting olives in the case of the 72 regular images, as it exactly segmented 50 olives in all of them (3,600 in total), independently from the olive variety considered. With respect to the 45 stress images, they represented a very more complex and challenging scenario for the algorithm, as they contained numerous olives chaotically positioned, with multiple border cuts and olive touching occurrences to deal with.

Fig. 12 illustrates a set of stress-images and the graphical representation of the results of their analysis.

As it can be observed, the algorithm showed great ability to individually segment olives and to discard those cut by the image borders. This visual impression was corroborated by the numerical analysis of the results given in Table 2. This table compiles the results regarding fruit-detection

TABLE 3. External validation results calculated considering all varieties as a whole (N = 2,700).

Feature	Model type	RMSE	SE (%)	|E| (%)
Major axis	Variety-dependent	0.4956 (mm)	2.67	0.44
	Variety-independent	0.5714 (mm)	3.07	1.14
Minor axis	Variety-dependent	0.3543 (mm)	2.47	0.34
	Variety-independent	0.3711 (mm)	2.57	0.60
Mass: $Area$	Variety-dependent	0.1192 (g)	5.20	0.55
	Variety-independent	0.2106 (g)	9.89	6.46
Mass: $Area_{NLog}$	Variety-dependent	0.1134 (g)	4.84	0.51
	Variety-independent	0.2806 (g)	13.53	9.05
Mass: $Area_{Log_2}$	Variety-dependent	0.1135 (g)	4.84	0.50
	Variety-independent	0.2878 (g)	13.87	9.39
Mass: $Area_{Log_3}$	**Variety-dependent**	**0.1132 (g)**	**4.84**	**0.49**
	Variety-independent	0.2783 (g)	13.42	8.95
Mass: $Area_{Log_5}$	Variety-dependent	0.1132 (g)	4.85	0.49
	Variety-independent	0.2675 (g)	12.89	8.43
Mass: $Area_{Log_{10}}$	Variety-dependent	0.1139 (g)	4.88	0.54
	Variety-independent	0.2550 (g)	12.27	7.76
Mass: $Area_{Root_2}$	Variety-dependent	0.1141 (g)	4.89	0.56
	Variety-independent	0.2488 (g)	11.94	7.38
Mass: $Area_{Root_3}$	Variety-dependent	0.1147 (g)	4.95	0.50
	Variety-independent	0.2329 (g)	11.12	6.47
Mass: $Area_{Root_5}$	Variety-dependent	0.1161 (g)	5.03	0.51
	Variety-independent	0.2222 (g)	10.55	6.40
Mass: $Area_{Root_7}$	Variety-dependent	0.1170 (g)	5.08	0.54
	Variety-independent	0.2183 (g)	10.34	6.36
Mass: $Area_{Root_{10}}$	Variety-dependent	0.1176 (g)	5.11	0.54
	Variety-independent	0.2157 (g)	10.19	6.38

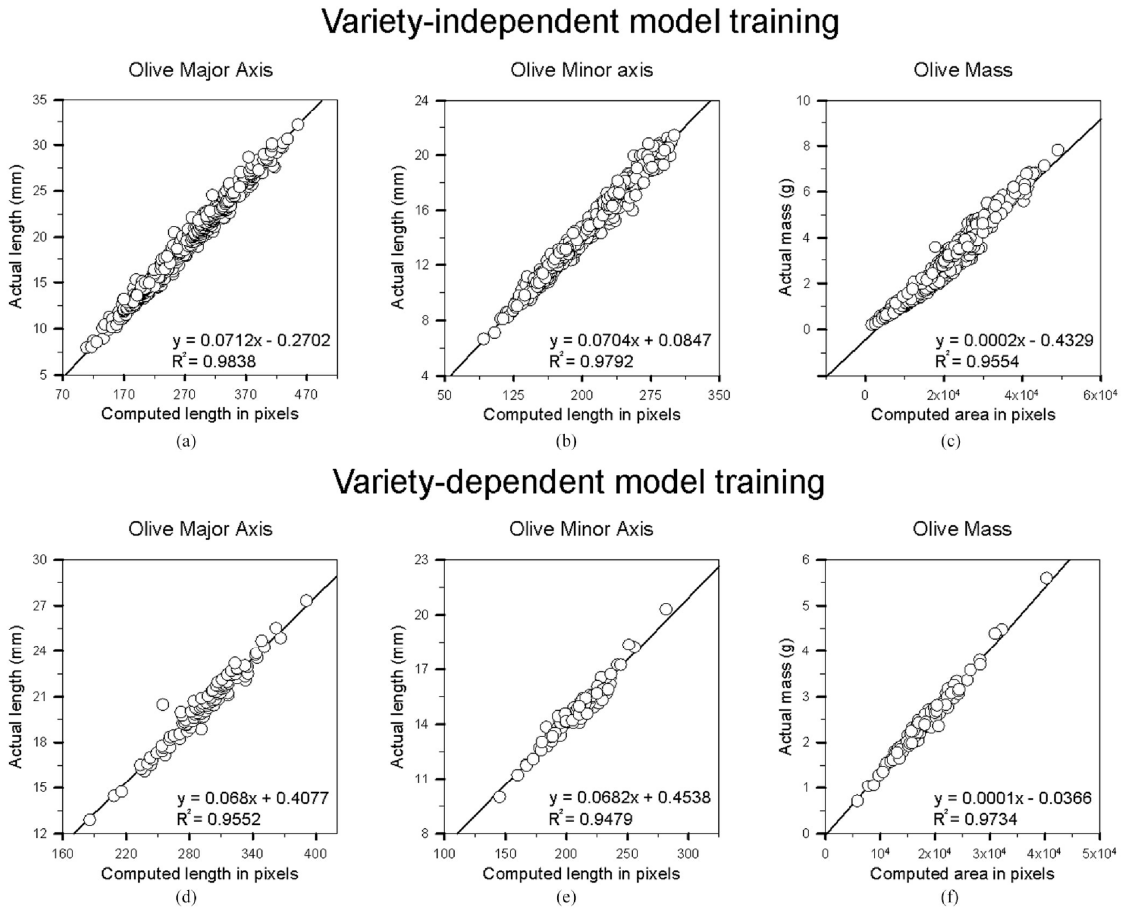


FIGURE 13. Representation of the regression analysis performed to compute the different variety-independent and -dependent models: (a) major-axis-length variety-independent model; (b) minor-axis-length variety-independent model; (c) $Area_{Log_3}$ mass variety-independent model; (d) major-axis-length model for changlot real variety; (e) minor-axis-length model for changlot real variety; (f) $Area_{Log_3}$ mass model for changlot real variety.

accuracy, by showing the ratio between the olives detected and the actual number of fruits which appear in the corresponding stress images.

An outstanding number of 8,106 olive instances were present in the 45 stress images. The algorithm individually segmented 8,110. This slight deviation responds to the overestimation of 5 olives in 4 different stress images, all of them related to not correctly removing olives cut by the image border. Additionally, there was only one case of underestimation in which a pair of fruits, touching each other, were not properly separated; it occurred in image *stress_image_2* of the Picual variety. Thus, if 6 segmentation errors are counted over 8106 cases, it can be concluded that the accuracy of the algorithm measured on the stress images was 99.92%.

B. RESULTS OF THE WHOLE METHODOLOGY FOR ESTIMATING OLIVE-FRUITS MASS AND SIZE

A set of 27 variety-dependent models (9 varieties × 3 features), and a set of 3 variety-independent models (1 per feature for all varieties), were obtained by linearly correlating the features obtained by image analysis to the actual reference measures. Overall, those mass estimation models trained featuring olive mass with the weighing function $Area_{Log_3}$ (eq. (23), $n = 3$), provided better performance at external validation than most of the other approaches. Table 3 summarizes the results of this comparison computed on the external validation sets presenting, per model tentative, the mean value of the previously defined error metrics calculated for each olive variety. Consequently, the discussion and the results reviewed hereafter regarding mass estimation are referred to this choice.

Fig. 13 illustrates the correlation results registered when training the three independent models and, as an example, the three dependent models for Changlot Real variety. As it can be observed, there was a significant positive correlation, exhibiting a minimal degree of dispersion in all the cases. This fact, extendable to the rest of the models computed, highlights a strong statistical relationship between the

TABLE 4. Training and external validation results for each of the estimation models proposed and varieties studied. N = 900 for model training (100 per variety), and N = 2,700 for validation (300 per variety).

			TRAINING		EXTERNAL VALIDATION		
Variety	Feature	Model type	R^2	Regression equation	*RMSE*	*SE* (%)	$\|E\|$ (%)
967	Major axis	Variety-dependent	0.9407	y=0.06427x+0.95664	0.3946	2.64	0.68
		Variety-independent	0.9838	y=0.07124x-0.27022	0.4242	2.84	1.17
	Minor axis	Variety-dependent	0.9544	y=0.06724x+0.47133	0.3708	2.95	1.04
		Variety-independent	0.9792	y=0.07037x+0.08466	0.3474	2.76	0.32
	Mass: $Area_{Log_3}$	Variety-dependent	0.9756	y=0.00012x-0.03778	0.0715	5.07	0.63
		Variety-independent	0.9554	y=0.00016x-0.43286	0.1652	11.71	5.16
1030	Major axis	Variety-dependent	0.9439	y=0.06515x+1.01841	0.3478	2.42	0.53
		Variety-independent	0.9838	y=0.07124x-0.27022	0.3862	2.69	0.86
	Minor axis	Variety-dependent	0.9652	y=0.06913x+0.33592	0.2528	2.16	0.02
		Variety-independent	0.9792	y=0.07037x+0.08466	0.2584	2.21	0.41
	Mass: $Area_{Log_3}$	Variety-dependent	0.9836	y= 0.00014x-0.03032	0.0456	3.88	0.15
		Variety-independent	0.9554	y=0.00016x-0.43286	0.1976	16.79	15.19
Arbequina	Major axis	Variety-dependent	0.9772	y= 0.06448x+0.9561	0.5037	3.42	0.58
		Variety-independent	0.9838	y=0.07124x-0.27022	0.5710	3.88	0.83
	Minor axis	Variety-dependent	0.9747	y= 0.06771x+0.51259	0.2889	2.17	0.17
		Variety-independent	0.9792	y=0.07037x+0.08466	0.3052	2.29	0.40
	Mass: $Area_{Log_3}$	Variety-dependent	0.9886	y= 0.00013x+0.02085	0.1106	7.11	0.60
		Variety-independent	0.9554	y=0.00016x-0.43286	0.2109	13.57	7.65
Arbosana	Major axis	Variety-dependent	0.9645	y= 0.06933x+0.04436	0.3154	2.22	0.11
		Variety-independent	0.9838	y=0.07124x-0.27022	0.3321	2.34	0.41
	Minor axis	Variety-dependent	0.9471	y=0.06897x+0.29699	0.2260	2.01	0.18
		Variety-independent	0.9792	y=0.07037x+0.08466	0.2336	2.08	0.28
	Mass: $Area_{Log_3}$	Variety-dependent	0.9816	y=0.00011x+0.00384	0.0435	4.02	0.34
		Variety-independent	0.9554	y=0.00016x-0.43286	0.1844	17.01	6.87
Changlot Real	Major axis	Variety-dependent	0.9552	y=0.06801x+0.4077	0.5938	2.89	0.78
		Variety-independent	0.9838	y=0.07124x-0.27022	0.5970	2.91	0.54
	Minor axis	Variety-dependent	0.9479	y=0.06824x+0.45384	0.3035	2.06	0.45
		Variety-independent	0.9792	y=0.07037x+0.08466	0.2988	2.03	0.06
	Mass: $Area_{Log_3}$	Variety-dependent	0.9734	y=0.00014x-0.03664	0.1056	4.25	1.16
		Variety-independent	0.9554	y=0.00016x-0.43286	0.1601	6.44	0.70
Lechín de Sevilla	Major axis	Variety-dependent	0.9451	y=0.0734x-0.76493	0.5755	2.76	0.24
		Variety-independent	0.9838	y=0.07124x-0.27022	0.5913	2.84	0.92
	Minor axis	Variety-dependent	0.8791	y=0.06498x+1.13778	0.3418	2.30	0.27
		Variety-independent	0.9792	y=0.07037x+0.08466	0.3755	2.53	0.86
	Mass: $Area_{Log_3}$	Variety-dependent	0.9673	y=0.00012x+0.00431	0.1059	4.15	0.28
		Variety-independent	0.9554	y=0.00016x-0.43286	0.4007	15.71	12.46
Ocal	Major axis	Variety-dependent	0.9282	y=0.06567x+1.93467	0.7471	2.89	0.64
		Variety-independent	0.9838	y=0.07124x-0.27022	0.8365	3.24	1.37
	Minor axis	Variety-dependent	0.8463	y=0.06721x+1.0975	0.5275	2.79	0.38
		Variety-independent	0.9792	y=0.07037x+0.08466	0.5980	3.16	1.32
	Mass: $Area_{Log_3}$	Variety-dependent	0.9596	y=0.00015x+0.22991	0.2986	5.91	0.85
		Variety-independent	0.9554	y=0.00016x-0.43286	0.5747	11.37	9.68
Picual	Major axis	Variety-dependent	0.9486	y=0.0692x+0.87548	0.4525	2.32	0.20
		Variety-independent	0.9838	y=0.07124x-0.27022	0.7827	4.01	3.25
	Minor axis	Variety-dependent	0.9406	y=0.06963x+0.37024	0.2990	2.32	0.14
		Variety-independent	0.9792	y=0.07037x+0.08466	0.3276	2.55	1.03
	Mass: $Area_{Log_3}$	Variety-dependent	0.9773	y=0.00012x-0.07138	0.0749	4.17	0.27
		Variety-independent	0.9554	y=0.00016x-0.43286	0.3767	20.97	18.42
Verdial de Huevar	Major axis	Variety-dependent	0.9492	y=0.06607x+1.14255	0.5299	2.50	0.20
		Variety-independent	0.9838	y=0.07124x-0.27022	0.6221	2.94	0.94
	Minor axis	Variety-dependent	0.8506	y= 0.06252x+1.98224	0.5783	3.47	0.40
		Variety-independent	0.9792	y=0.07037x+0.08466	0.5956	3.58	0.74
	Mass: $Area_{Log_3}$	Variety-dependent	0.9658	y= 0.00015x+0.00958	0.1622	4.97	0.11
		Variety-independent	0.9554	y=0.00016x-0.43286	0.2348	7.20	4.44

estimations and actual values, which reinforce the viability of using simple-linear-regression-based modelling in order to individually estimate the mass and size of olive-fruits.

Table 4 details training and external validation results for all the model variants and olive varieties. With regard to training, outstandingly high values were registered for the coefficient of determination, R^2, in all cases. Indeed, the coefficient value was below 0.9 only in three cases, reaching 0.8463 in the worst case. Furthermore, training was generally so consistent and comparable for both the variety-dependent and the variety-independent approaches, that no conclusions could be taken at this stage. However, when applying the

TABLE 5. Results of one-way analysis of variance performed on the estimations produced by the variety-dependent and -independent models for each variety. The analyzed populations are the individual relative errors, calculated by using (34), produced by the different models on the external validation sets of each variety (N = 2,700, 300 per variety). The mean $\overline{e_i}$ and standard deviation σ_{e_i} of each population is given. dissimilar letters indicate different statistical means according to the analysis of variance using the tukey's test at $p < 0.05$.

Variety	Variety-dependent models			Variety-independent models		
	Major axis $(\overline{e_i}, \sigma_{e_i})$	Minor axis $(\overline{e_i}, \sigma_{e_i})$	Mass: $Area_{Log_3}$ $(\overline{e_i}, \sigma_{e_i})$	Major axis $(\overline{e_i}, \sigma_{e_i})$	Minor axis $(\overline{e_i}, \sigma_{e_i})$	Mass: $Area_{Log_3}$ $(\overline{e_i}, \sigma_{e_i})$
967	(0.9944,0.0265) [a]	(0.9907,0.0276) [a]	(0.9964,0.0488) [ab]	(1.0117,0.0268) [d]	(1.0039,0.0279) [de]	(1.0225,0.1314) [d]
1030	(0.9952,0.0243) [a]	(1.0003,0.0218) [bc]	(1.0020,0.0468) [b]	(0.9909,0.0264) [b]	(0.9961,0.0219) [bc]	(0.8148,0.1503) [a]
Arbequina	(0.9950,0.0339) [a]	(0.9985,0.0223) [bc]	(0.9974,0.0758) [ab]	(1.0071,0.0377) [cd]	(1.0037,0.0232) [de]	(0.8770,0.1725) [b]
Arbosana	(0.9989,0.0234) [ab]	(1.0018,0.0204) [bc]	(0.9990,0.0418) [ab]	(1.0037,0.0242) [c]	(1.0026,0.0210) [de]	(1.0131,0.2182) [d]
Changlot Real	(0.9927,0.0280) [a]	(0.9959,0.0202) [ab]	(0.9884,0.0423) [a]	(1.0054,0.0290) [cd]	(1.0008,0.0205) [cd]	(0.9941,0.0723) [d]
Lechín de Sevilla	(0.9973,0.0287) [ab]	(1.0034,0.0231) [c]	(0.9979,0.0438) [ab]	(0.9910,0.0277) [b]	(1.0085,0.0240) [e]	(1.1065,0.0916) [e]
Ocal	(0.9944,0.0282) [a]	(0.9964,0.0278) [ab]	(0.9948,0.0598) [ab]	(0.9861,0.0297) [b]	(0.9867,0.0290) [a]	(0.8983,0.0648) [b]
Picual	(0.9984,0.0235) [ab]	(1.0018,0.0235) [bc]	(1.0033,0.0436) [b]	(0.9673,0.0240) [a]	(0.9899,0.0234) [ab]	(1.1714,0.0858) [f]
Verdial de Huévar	(1.0022,0.0265) [b]	(0.9975,0.0340) [bc]	(1.0005,0.0501) [ab]	(1.0086,0.0296) [cd]	(0.9930,0.0345) [ab]	(0.9430,0.0755) [c]

Dissimilar Letters Indicate Different Statistical Means According to the Analysis of Variance Using the Tukey's Test at p<0.05

trained models to the sets of external validation, the calculation of the error metrics to assess the given different estimations showed some certain trends to be analyzed.

On the one hand, focusing on mass estimation, variety-dependent models offered, systematically and by far, lower estimation error rates than the independent ones. Actually, variety-dependent models gave $|E|$ rates per variety for mass estimation below 1% with the exception of that for Changlot Real, which reached 1.16%; the mean error $|E|$ considering all varieties as a whole was 0.49% (Table 3). The trend is also corroborated by the results registered with metrics based on accumulated error, *RMSE* and *SE*. The higher error rates for the independent approach can be explained by differences in physical characteristic features, as fruit density, stone-fresh ratio, or fat content, among others. All of them, having an evident impact on mass, may differ from one variety to another. Consequently, the slope of the lines better fitting the mass-size relationship may vary among varieties, thus explaining the poorer behavior shown by the independent approach. Studied this effect, no definitive reasons can be given to discard generic variety-independent models as a feasible tool for mass estimation, but it suggests the necessity to explore alternative approaches for its more effective implementation. To this end, a future study will explore non-linear modelling, in an attempt to achieve models able to express this mass-size variability.

Respecting size estimation, satisfying results were also found. As a reference, the mean error $|E|$ for the major and minor axis length estimation was lower than 1%, excepting for a few cases, per variety and considering all of them together (Table 4 and Table 3, respectively). Notwithstanding, despite a slightly more accurate behavior could be recognized for the variety-dependent models, this difference was so subtle that suggested further analysis in order to be more conclusive. Table 5 shows the results of the one-way analysis of variance performed per variety and model approach.

TABLE 6. Results of the study of variability of manual olive size measurement performed by three different observers. N = 900, 100 per variety, where each instance is the deviation of the three individual observations given by (36).

Variety	Major Axis		Minor Axis	
	Mean Length (mm)	SME (%)	Mean Length (mm)	SME (%)
967	15.27	1.21	13.03	1.19
1030	14.59	1.19	11.92	1.30
Arbequina	15.10	2.21	13.64	0.94
Arbosana	14.32	0.93	11.31	1.0
Changlot	20.81	0.67	14.94	0.72
Lechín	20.03	1.05	14.35	1.36
Ocal	26.00	1.38	18.96	1.96
Picual	19.68	1.06	12.93	1.27
Verdial	20.67	0.94	16.38	1.08

Generally, a higher number of classes of significance resulted from the analysis of the estimations produced by the variety-independent models for the three features. Especially relevant was this difference for the case of mass, with 2 Vs 6 classes of significance, which confirms the previous conclusions and adds strong signs of statistical consistence regarding the behavior of this set of variety-dependent models. Contrary, differences were less conclusive for the case of size estimation. This fact encouraged to study the reliability of the reference objective measurements, for the major and minor axis length, taken in laboratory by hand using a digital caliper. Indeed, there are multiple factors which potentially introduce uncertainty in the procedure, as the slight elasticity of the fruits, their ellipsoidal shape and other morphological features. Table 6 compiles the results of the variability of the manual measurements performed by three different observers for a set of 900 individuals, 100 per variety. As it can be observed, the magnitude of the found variability is comparable to the analyzed slight performance deviation between the variety-dependent and variety-independent models. As this

variability may affect model training and validation in a rather random manner, the most conservative conclusion is to consider both modelling approaches comparably valid for the case of size estimation.

V. CONCLUSIONS

The present paper proposes a new methodology for the automatic counting, and the individual mass and size estimation, of olive-fruits in digital images. All this work is intended to assess the viability of a machine vision system, aimed at improving post-harvest olive-fruit grading and classification, currently performed by rather limited mechanical methods, and to lay the basis for its implementation. Thus, the proposed imaging chamber, designed to be potentially integrated in real classification machines equipped with translucent belts, provided an efficient way to carry out the acquisition of the digital images. On the other hand, the image analysis algorithm was conceived to deal with complex situations as fruit touching, or fruit cut by the image borders. It argues for the methodology could operate under a real scenario. Regarding estimation models, and by relying on external validation results, variety-dependent models outperformed the independent ones when estimating olive mass. It indicates mass-size ratio dependence from variety. Variety-dependent models showed slightly better behavior than the variety-independent ones for size estimation, again in terms of external validation. Notwithstanding, as variability of the manual reference measurements was found to have a magnitude comparable to the performance deviation of both approaches, the most conservative conclusion is to consider them analogously valid.

A. FUTHER WORK

As detailed before, 3,600 fruits from nine different olive varieties where used to train and validate the methodology. Despite these figures, along with the obtained results, support confidence in the developed methodology, further research will include new olive-fruit varieties and samples, in order to reinforce its generalization.

On the other hand, variety-dependent models consistently and outstandingly outperformed the independent ones for the case of mass estimation, providing an overall mean error |E| of 0.49%. It indicates mass-size relationship dependence from olive variety, what encourages to explore in the future non-linear modelling for the case of independent methods, in an attempt to achieve statistical models able to express this mass-size variability.

As stated before, the image analysis algorithm was developed by using MATLAB along with the Image Processing Toolbox for prototyping purposes. Future research will also be focused on the implementation of the methodology within a framework able to satisfy real-time computation.

ACKNOWLEDGMENT

The authors would like to thank ''Cooperativa Virgen de la Oliva'', for generously offering their orchards to conduct this work.

REFERENCES

[1] P. Vossen, ''Olive oil: History, production, and characteristics of the world's classic oils,'' *Amer. Soc. Horticultural Sci.*, vol. 42, no. 5, pp. 1093–1100, 2007.

[2] J. Wang, D. Zhang, T. J. A. Farooqi, L. Ma, Y. Deng, and Z. Jia, ''The olive (Olea europaea L.) industry in China: Its status, opportunities and challenges,'' *Agrofor. Syst.*, vol. 93, no. 2, pp. 395–417, 2019.

[3] International Olive Council (IOC). (2018). *World Statistics on Production, Imports, Exports and Consumption. Word Table Olive Figures. Production*. Accessed: Jan. 8, 2019. [Online]. Available: http://www.internationaloliveoil.org/estaticos/view/132-world-table-olive-figures

[4] International Olive Council (IOC). (2018). *World Statistics on Production, Imports, Exports and Consumption. Word Olive Oil Figures. Production*. Accessed: Jan. 8, 2019. [Online]. Available: http://www.internationaloliveoil.org/estaticos/view/131-world-olive-oil-figures

[5] A. Scheidel and F. Krausmann, ''Diet, trade and land use: A socio-ecological analysis of the transformation of the olive oil system,'' *Land Policy*, vol. 28, no. 1, pp. 47–56, Jan. 2011.

[6] G. Giametta and B. Bernardi, ''Olive grove equipment technology. Straddling trees: Mechanized olive harvests,'' *Adv. Horticultural Sci.*, vol. 24, no. 1, pp. 64–70, 2010.

[7] A. H. S. Gómez, P. G. García, and L. R. Navarro, ''Elaboration of table olives,'' *Grasas y Aceites*, vol. 57, no. 1, pp. 86–94, 2006.

[8] S. Kailis and D. Harris, ''Quality and safety evaluation of processed olives,'' in *Producing Table Olives*, Collingwood, VIC, Australia: Landlinks Press, 2007, ch. 6, pp. 244–246.

[9] G. P. Moreda, J. Ortiz-Cañavate, F. J. García-Ramos, and M. Ruiz-Altisent, ''Non-destructive technologies for fruit and vegetable size determination—A review,'' *J. Food Eng.*, vol. 92, no. 2, pp. 119–136, May 2009.

[10] F. Jiménez-Jiménez, S. Castro-García, G. L. Blanco-Roldán, L. Ferguson, U. A. Rosa, and J. A. Gil-Ribes, ''Table olive cultivar susceptibility to impact bruising,'' *Postharvest Biol. Technol.*, vol. 86, pp. 100–106, Dec. 2013.

[11] V. Van Linden, N. Scheerlinck, M. Desmet, and J. De Baerdemaeker, ''Factors that affect tomato bruise development as a result of mechanical impact,'' *Postharvest Biol. Technol.*, vol. 42, no. 3, pp. 260–270, Dec. 2006.

[12] N. Kondo, ''Automation on fruit and vegetable grading system and food traceability,'' *Trends Food Sci. Technol.*, vol. 21, no. 3, pp. 145–152, Mar. 2010.

[13] M. Baigvand, A. Banakar, S. Minaei, J. Khodaei, and N. Behroozi-Khazaei, ''Machine vision system for grading of dried figs,'' *Comput. Electron. Agricult.*, vol. 119, pp. 158–165, Nov. 2015.

[14] F. S. A. Sa'ad, M. F. Ibrahim, A. Y. M. Shakaff, A. Zakaria, and M. Z. Abdullah, ''Shape and weight grading of mangoes using visible imaging,'' *Comput. Electron. Agricult.*, vol. 115, pp. 51–56, Jul. 2015.

[15] Z. Wang, K. B. Walsh, and B. Verma, ''On-tree mango fruit size estimation using RGB-D images,'' *Sensors*, vol. 17, no. 12, p. 2738, Nov. 2017.

[16] A. Mizushima and R. Lu, ''An image segmentation method for apple sorting and grading using support vector machine and Otsu's method,'' *Comput. Electron. Agricult.*, vol. 94, pp. 29–37, Jun. 2013.

[17] G. Gatica, S. Best, J. Ceroni, and G. Lefranc, ''Olive fruits recognition using neural networks,'' *Procedia Comput. Sci.*, vol. 17, pp. 412–419, Jan. 2013.

[18] S. S. Martínez, D. M. Gila, A. Beyaz, J. G. Ortega, and J. G. García, ''A computer vision approach based on endocarp features for the identification of olive cultivars,'' *Comput. Electron. Agricult.*, vol. 154, pp. 341–346, Nov. 2018.

[19] A. Beyaz, M. T. özkaya, and D. Içen, ''Identification of some spanish olive cultivars using image processing techniques,'' *Sci. Horticult.*, vol. 225, pp. 286–292, Nov. 2017.

[20] R. Diaz, L. Gil, C. Serrano, M. Blasco, E. Moltó, and J. Blasco, ''Comparison of three algorithms in the classification of table olives by means of computer vision,'' *J. Food Eng.*, vol. 61, no. 1, pp. 101–107, Jan. 2004.

[21] D. A. Puerto, D. M. M. Gila, J. G. García, and J. G. Ortega, ''Sorting olive batches for the milling process using image processing,'' *Sensors*, vol. 15, no. 7, pp. 15738–15754, Jul. 2015.

[22] J. M. Ponce, A. Aquino, B. Millán, and J. M. Andújar, ''Olive-fruit mass and size estimation using image analysis and feature modeling,'' *Sensors*, vol. 18, no. 9, p. 2930, Sep. 2018.

[23] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Germany, Berlin: Springer, 2003.

[24] M. Sonka, V. Hlavac, and R. Boyle, "Color spaces," in *Image Processing, Analysis, and Machine Vision*, 4th ed. Belmont, CA, USA: Cengage Learning, 2014, pp. 36–38.
[25] W. Castro, J. Oblitas, M. De-la-Torre, C. Cotrina, K. Bazán, and H. Avila-George, "Classification of Cape gooseberry fruit according to its level of ripeness using machine learning techniques and different color spaces," *IEEE Access*, vol. 7, pp. 27389–27400, 2019.
[26] F. Mendoza, P. Dejmek, and J. M. Aguilera, "Calibrated color measurements of agricultural foods using image analysis," *Postharvest Biol. Technol.*, vol. 41, no. 3, pp. 285–295, 2006.
[27] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[28] S. Beucher and C. Lantuéjoul, "Use of watersheds in contour detection," in *Proc. Int. Workshop Image Process. Real-Time Edge Motion Detection*, Rennes, France, 1979, pp. 391–396.
[29] S. Beucher and F. Meyer, "The morphological approach to segmentation: The watershed transformation. Mathematical morphology in image processing," *Opt. Eng.*, vol. 34, pp. 433–481, 1993.
[30] F. Meyer and S. Beucher, "Morphological segmentation," *J. Vis. Commun. Image Represent.*, vol. 1, no. 1, pp. 21–46, 1990.
[31] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.
[32] A. K. Jain, "Image analysis and computer vision," in *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989, ch. 9, pp. 377–394.
[33] J. W. Tukey, "Comparing individual means in the analysis of variance," *Biometrics*, vol. 5, no. 2, pp. 99–114, Jun. 1949.

JUAN M. PONCE was born in Huelva, Andalusia, Spain, in 1982. He received the Engineer's degree in computer sciences from the University of Seville, Seville, in 2009.

After performing different jobs within the software development industry as a Programmer and Analyst, nowadays, he is with the University of Huelva, Huelva, where he is currently pursuing the Ph.D. degree in image analysis. His research is focused on developing computer-vision-based technology, potentially applicable in the food industry and precision agriculture.

ARTURO AQUINO was born in Huelva, Andalusia, Spain, in 1982. He received the M.Sc. degree in computer engineering and networks from the University of Granada, in 2007, and the Ph.D. degree in image analysis from the University of Huelva, in 2011.

He was an Engineer in computer sciences with the University of Seville, in 2006. Since 2007, he has been working in various academic posts at the University of Granada; the University of Huelva; the Research Centre for Mathematical Morphology, Mines ParisTech; and the University of La Rioja. He is currently a Postdoctoral Researcher and Lecturer with the University of Huelva. Until 2019, he has published over 40 papers, including articles, book chapters, and contributions to conferences, and holds two patents. He has directed a doctoral thesis and has participated in eight European, national, and regional research projects. His research interests include image analysis, processing, and understanding.

BORJA MILLAN was born in Zaragoza, Spain, in 1982. He received the B.S. degree in industrial engineering and the Ph.D. degree in sustainable agricultural ecosystems from the University of La Rioja, Logroño, in 2011 and 2017, respectively.

He is currently teaching and researching with the Department of Electronic Engineering, Computer Systems and Automation, University of Huelva, with a Juan de la Cierva Grant from the Spanish Ministry of Science, Innovation, and Universities. He has published over 15 publications indexed in JCR, five patents, and has worked in five European, four national, and four regional projects. He has carried out research stays at the Valencian Institute of Agrarian Research (IVIA) and at the Center for Mathematical Morphology (CMM) of the University of ParisTech. His research interests include computer vision, remote sensing, and electronics applied to precision agriculture.

Dr. Millan has received several awards and honors, including the first prize for final career projects from the College of Industrial Engineers of Aragon and La Rioja and obtained the first prize in the Business Ideas Competition from the Chair of Entrepreneurs at the University of La Rioja.

JOSÉ M. ANDÚJAR was born in Huelva, Spain. He received the Ph.D. degree from the University of Huelva, in 2000. He is currently a Full Professor of systems engineering and automatic control with the University of Huelva, Spain. He has conducted 10 Doctoral Theses with eight prizes, and holds 12 international patents. He has over 100 papers published in indexed journals in the ISI Journal Citation Reports. Specifically, he has 51 quartile Q1 publications in 20 different journals; most of these are among the top 10 in their categories, and several are number one. He has led or co-led over 50 research projects funded by public institutions and companies. His main research interests are control engineering, renewable energy systems, remote piloted aircraft systems applications, and engineering education. Throughout his professional life, he has received 24 awards and academic honors.

4.1.3. Article 3

Olive-Fruit Variety Classification by Means of Image Processing and Convolutional Neural Networks

J. M. Ponce, A. Aquino, J. M. Andújar

Published in:

Journal: IEEE Access (ISSN: 2169-3536)

Editorial: Institute of Electrical and Electronics Engineers (IEEE)

Reference: vol. 7, pp. 147629-147641

Year: 2019

DOI: 10.1109/ACCESS.2019.2947160

Quality index (Journal Citation Reports®, 2019): 26/90 (Q2) in the category "Telecommunications", 35/156 (Q1) in "Computer Science, Information Systems" and 61/266 (Q1) in "Engineering, Electrical & Electronic". Impact Factor of 3.745.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 23, 2019, accepted October 8, 2019, date of publication October 14, 2019, date of current version October 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2947160

Olive-Fruit Variety Classification by Means of Image Processing and Convolutional Neural Networks

JUAN M. PONCE, ARTURO AQUINO, AND JOSÉ M. ANDÚJAR, (Senior Member, IEEE)
Department of Electronic Engineering, Computer Systems and Automation, Higher Technical School of Engineering, University of Huelva, 21007 Huelva, Spain
Corresponding author: Juan M. Ponce (jmponce.real@diesia.uhu.es)

This work was supported in part by the INTERREG Cooperation Program V-A SPAIN-PORTUGAL (POCTEP) 2014–2020, and in part by the ERDF funds within the scope of the TecnOlivo Project under Grant 0155_TECNOLIVO_6_E.

ABSTRACT The automation of classification and grading of horticultural products attending to different features comprises a major challenge in food industry. Thus, focused on the olive sector, which boasts of a huge range of cultivars, it is proposed a methodology for olive-fruit variety classification, approaching it as an image classification problem. To that purpose, 2,800 fruits belonging to seven different olive varieties were photographed. After processing these initial captures by means of image processing techniques, the resulting set of images of individual fruits were used to train, and continuedly to externally validate, the implementations of six different Convolutional Neural Networks architectures. This, in order to compute the classifiers with which perform the variety categorization of the fruits. Remarkable hit rates were obtained after testing the classifiers on the corresponding external validation sets. Thus, it was yielded a top accuracy of 95.91% when using the Inception-ResnetV2 architecture. The results suggest that the proposed methodology, once integrated into industrial conveyor belts, promises to be an advanced solution to postharvest olive-fruit processing and classification.

INDEX TERMS Computer vision, convolutional neural network, fruit variety, food industry, fruit classification, image processing, olive.

I. INTRODUCTION

Olive (Olea europaea L.) growing is currently an agronomic activity of great importance. With an ancient tradition throughout the Mediterranean basin, its cultivation has spread around the world in recent decades [1], [2], and consumption of table olives and olive oil, which are the most important products derived from this crop, have exploded. Indeed, in accordance with the International Olive Council (IOC), the consumption of table olives has more than doubled in the past 20 years [3], with an estimation of 2,667,000 tonnes in 2018/19 against the 1,185,500 tonnes consumed worldwide during the year 1998/1999. In the case of olive oil, the numbers are equally significant, and in the same 20-year period its consumption has increased in more than 500,000 tonnes, being estimated 2,950,500 tons for the year 2018-19 [4].

Therefore, and as it is happening with other mainstream crops, the olive sector has to face multiple challenges in order to satisfy this high demand market, in which the popularity of olive-derived products does not stop growing. Increasing production while reducing the associated costs, and all this in an environmentally sustainable way, is a cross-cutting problem in current agriculture [5]–[7]. The introduction of new technologies is playing a fundamental role to deal with this situation. And this is happening in virtually every scope related to agricultural activities. In this sense, postharvest tasks have become an important spotlight. Indeed, optimising processes involved in the treatment and manipulation of horticultural products, once they are gathered, may have a remarkable cost-saving impact [8].

Within this context, postharvest classification of horticultural products, according to different features such as size or surface condition, has become a main focus of research. This, since it has been traditionally performed manually, implying tedious, inaccurate and time-consuming tasks. Hence, its automation has become a major problem in food industry. The case of olive sector is not an exception, as postharvest fruit classification remains a challenge for olive growers.

The associate editor coordinating the review of this manuscript and approving it for publication was Gianluigi Ciocca.

Thus, while the automation of the olive-fruit size grading has historically found different solutions, all of them usually based on mechanical approaches, the classification according different criteria, with a potential value for farmers and producers, is still attached to manual inspection. This may be the case of individually classifying olive fruits according to their variety, feature in which the present study is focused on. By not discriminating among varieties when gathering the olive-fruits, harvest costs can be potentially reduced. It should be noted that harvesting is a key factor for olive growing, obviously directly related to the prize of the eventual product, and with a critical impact in the productivity and viability of olive growing as a business. Therefore, olive sector has put huge efforts to improve and optimise harvesting. New orchards planning or harvesting mechanization are some of the advances introduced during the last few decades in order to enhance fruit gathering [9], [10]. Indirectly, it may benefit from a postharvest automated classification system to separate the fruits, once they have been transported to the mill, according their variety. This could ease optimise harvesting, when orchards are shared by different varieties of olive-trees, not being necessary to consider this circumstance when planning the collection of the fruits.

Machine Vision systems, commonly used in a great variety of industries, have become familiar within the food sector, when inspecting and supporting the automatic handling of commodities [11]. By their integration throughout the manipulation process, the products to handle can be imaged, and descriptive features of each of them can be extracted in real time, via image processing and analysis. This information can be potentially used to label or categorize each of these products, according to models previously computed. So, this technology sets itself up as a potential solution to the automation of the olive-fruit variety-based classification.

In recent years, Computer Vision has significantly benefited from the recent revitalisation of Deep Learning, which has seen increased its popularity and it has experienced an exponential growth, both in terms of the amount of related research conducted and its applications. This sort of subset of machine learning methods is nothing new, and the principles it is based on have been around for a while [12], [13]. Nevertheless, the increase of computational power provided by graphics processing units (GPUs) [14], along with the huge amount of data offered by the Internet [15], have been seen as an opportunity to boost Deep Learning performance. Within this field, Convolutional Neural Networks (CNN) has supposed a breakthrough in image processing and analysis, and nowadays it comprises the main framework for image classifiers development and pattern recognition. Its applications have reached innumerable fields, ranging from the automotive industry, supporting self-driving cars [16]–[18], to the healthcare sector, with automatic diagnostic systems for the analysis of medical images [19], [20]. Smart farming and food sector are not an exception. Thus, over the past few years it has been published a considerable amount of literature related to the potential application of Deep Learning-based technologies to agriculture and food industry [21], [22]. Yield prediction [23]–[26], leaf defoliation estimation [27] or fruits and crops classification [11], [28]–[31] are just some examples of scopes within which the use of Deep Learning and CNNs has been explored.

Olive sector can potentially gain from the use of Deep Learning based-computer vision systems too. Hence, the main goal of this study is to assess the viability of performing a machine vision- and Deep Learning-based classification of olive-fruits according to their variety. To the best to authors' knowledge, the approach, methodology and results derived from this work are unseen in the literature to date. Whilst most research related to this topic draws on images of the endocarps of the fruits to that end [32]–[34], the methodology proposed here is focused on a non-invasive approach, in which the classification is carried out uniquely using the morphology of the olive-fruits as distinguishing feature. For that purpose, batches of fruits, stochastically disposed, were photographed inside an image acquisition chamber designed to be potentially integrable in current conveyor belts. Then, the initial captures were transformed by means of mathematical morphology and global thresholding techniques, obtaining a set of images of individual fruits. Next, different architectures of CNN were tried by training them with these images. Finally, the resulting classifiers were tested by using an external validation set to evaluate and compare their performance.

The present manuscript is organized as follows. First, section II overviews the fundamentals of CNNs as well as briefly describes the architectures of interest for this work. Section III-A specifies the olive varieties considered in the research, as well as different aspects related to fruit sample collection for configuring the data of study. Then, all aspects involved in the process of image acquisition are detailed (section III-B). Section III-C focuses on the framework employed to implement the proposed algorithms for image classification, emphasizing the different developing tools and technologies used. In section IV-A it is described the methodology to transform the initial captures into the individual fruit images that, as raw data, are used to try the different CNN architectures considered. CNN training and validation is detailed throughout sections IV-B-2 and IV-B-3, respectively. Section V exposes and discusses the results after validating the different image/olive-fruit-variety classifiers developed. Finally, section VI comprises the summary of the most relevant aspects concluded throughout the experimentation.

II. OVERVIEW OF CNN ARCHITECTURES

In the present study, olive-fruit classification is approached as an image classification problem, faced by means of machine learning, by using CNNs. Indeed, the irruption of Deep Learning as a mainstream technology has exploded in numerous and uneven domains [35], being Computer Vision a highly benefited niche [36], where CNNs have provided outstanding advances in the state of the art.

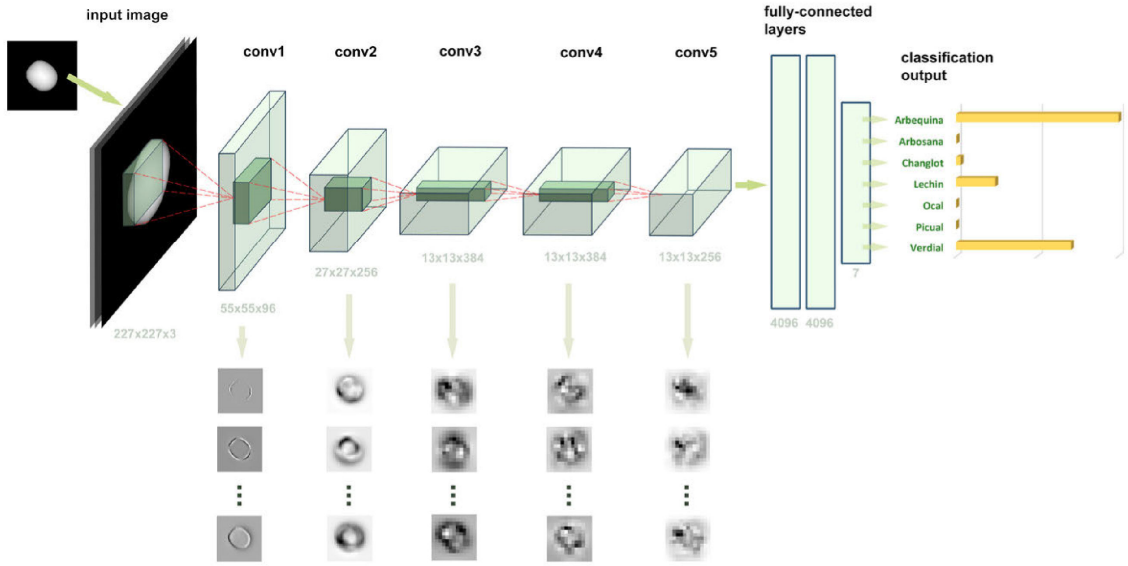


FIGURE 1. **Schematic illustration of the performance of a CNN, with five convolutional layers, in determining the variety of an olive fruit. The input is an olive-fruit image resulting from the pre-processing described in section IV-A. The image feeds the net and passes through the convolutional layers, which model the object from the gross distinctive features to the subtle ones as the image goes deeper. Finally, the extracted features are processed by a fully connected layer which yields a probability for the object to belong to the considered classes. In this example, the olive fruit is classified as *Arbequina*.**

Due to improvements in the design and manufacturing of specialized-in-parallel-computing integrated circuits, and the introduction of standardised architectures such as CUDA, GPUs have become an affordable and powerful framework for general purpose processing. This opened up a window of opportunity for researchers and engineers to face tasks that were not possible before because of the non-assumable computational cost. In addition, public datasets for machine learning purposes has increased in number over the past few years. Thus, different initiatives, such as ImageNet [37] or CIFAR-10 [38] among lots of others, have made available huge quantities of quality data. This without taking into account the exorbitant amount of information privately handled by the big players in the technology industry such as Amazon, Baidu, Facebook, Google or Microsoft, which far from being outsiders of this trend, have become in actual boosters of Deep Learning and its applications [15], [39].

Notwithstanding the computational improvements and data availability, classical approaches using neural networks dealt with the difficulty of finding and designing a set of significant mathematical descriptors [40] to feed the net with, in order to solve a specific problem. In contrast, the irruption of CNNs has supposed a game-changer in Computer Vision, as these networks can be directly fed with image-corresponding pixel-intensity matrixes. Thus, a CNN is basically a Deep Neural Network where, layer by layer, this raw data provided by the initial image, is transformed into a set of high-level features which are eventually used to realise the classification. To that end, CNN architectures are usually shaped by three different types of layers: convolutional layers in which filters or kernels, acting as feature detectors, are convolved over local regions of the input; pooling layers, where it is reduced the spatial dimensionality of the convolved features (also known as activation maps) obtained after a convolution stage; and fully-connected layers, with which, and by using the high-level features obtained by convolution and down-sampling, the class scores are calculated, and therefore the image classification is performed (see Fig. 1). Because of its demonstrated virtues, which include reduced complexity, faster model training, capacities of capturing local information, smaller sample volume requirements or lower overfitting probability, researchers have shown an increased interest, and CNNs applications and related literature [35], [36], [39], [41], that is encouraged to be consulted for deeper study, have exploded in recent years.

For the purpose of developing a set of models to undertake the olive-fruit variety classification, some of the most important CNN architectures seen in recent years are considered; they are briefly described hereafter.

AlexNet [42] supposed a breakthrough within the image recognition scope. It is based on an 8-layer architecture, consisting of 5 convolutional layers and 3 fully connected ones. So, deeper when compared to the standards of that time, another of its contributions was to model the output of the neurons in the network with a Rectified Linear Unit (ReLU), instead of other more conventional activation functions, like tanh or sigmoid. Thanks to that, the CNN could be trained much faster.

InceptionV1 [43], as well known as GoogLeNet, imposes 1×1 convolutions in the middle of the network, and proposes a global average pooling at the end of it, as a substitute for a last fully connected set of neurons as an output. Very deep when compared to AlexNet, the first iteration of the Inception architecture comprises 22 layers in total.

InceptionV3 [44] reduces the number of parameters in the network by factorizing convolutions. This, by increasing the number of layers (up to 42 in total) and reducing the dimensions of the filters, without loss of efficiency.

ResNet [45] implies a paradigm shift, introducing the notion of skip (or shortcut) connection. The layer inputs do not just depend on the outputs of the immediate previous layer. This, in order to solve a major problem with gradient-based learning methods, the vanishing/exploding gradients [46], [47]. Its architecture is complemented with a bottleneck design which allows to reduce complexity without significantly impoverishing performance. In this case, different alternatives can be found regarding the depth of the network: ResNet-50, with 50 layers; ResNet-101 with 101 layers; ResNet-152 with 152 layers.

Inception-ResNetV2 [48], based on the Inception architecture, but inspired by ResNet, is trained with residual connections, allowing to speed up the process.

III. MATERIALS AND METHODS

A. SAMPLE COLLECTION

Seven different olive varieties were considered for this study: *Arbequina*, *Arbosana*, *Picual*, *Ocal*, *Changlot Real*, *Verdial de Huévar* and *Lechín de Sevilla*. For each of these cultivars, 400 fruits were handpicked in olive orchards located in Gibraleón (37°20'09.2''N 7°02'19.8''W), province of Huelva (Andalusia, Spain), in October 2018.

B. IMAGE ACQUISITION

For the purpose of capturing the images of the fruits, it was set up an ad-hoc image acquisition system, inspired in that proposed in [49]. The aim at this point was not to implement a final version of this device, integrable in a conveyor system, but a prototype mimicking its main features regarding image capture conditions. This system was conceived to acquire images with high contrast in terms of luminosity between the fruits and the background, with a shadowless lighting. This feature favours a subsequent stage of the experimentation, where a solid and accurate segmentation of the captures is necessary.

The ad-hoc image acquisition system was designed to be potentially integrable in a conveyor belt. It consists of four main parts: a chamber with which the capture area can be isolated from any external light pollution; a lighting system, based on seven equally distributed strips of 25 5V-LEDs each, located at the base of the chamber; a semi-translucent white plastic surface comprising the capture area, over which olive-fruits are placed to be photographed; a digital mirrorless camera mounted above the chamber, looking perpendicularly at the capture area. Fig. 2 illustrates the described system.

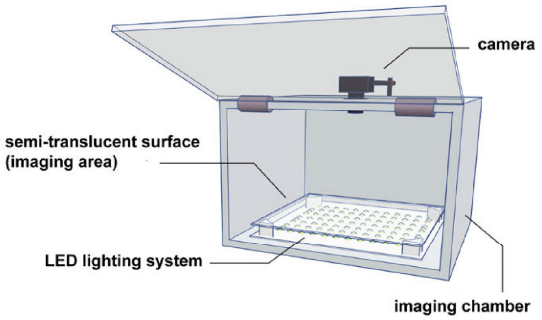


FIGURE 2. Image acquisition system.

FIGURE 3. Example of image acquired for Ocal variety.

It should be noted that the camera used for imaging the fruits, a Sony α7-II (Sony Corp., Tokyo, Japan), was equipped with a Zeiss 24/70mm lens (Carl Zeiss AG, Oberkochen, Germany) and configured in manual mode, with an aperture of f/7.1, shutter speed of 1/50s, focal length of 31mm and ISO sensitivity of 250. The initial captures were saved in JPEG format, with a resolution of 6000 $\times$ 3376 pixels and 24 bits of colour depth. The pixel density was set to 350ppi.

Every variety-set of 400 fruits was photographed in batches of 50 individuals, thus generating 8 images per variety. With the aim of mimicking a realistic scenario, it should be remarked that the fruits were stochastically placed on the capture area, with the only restriction of not appearing any of them touching the border in the resulting image. Fig. 3 shows an example of an image taken following the specified criteria.

C. IMAGE PROCESSING ALGORITHM IMPLEMENTATION AND CNN TRAINING/VALIDATION

The image processing algorithm, developed to transform the starting captures into the set of images to be used throughout the experimentation, was implemented with MATLAB and its Image Processing Toolbox, release 2018a (The MathWorks Inc., Natick, Massachusetts, USA). A MATLAB framework was also used in order to perform the training of the CNNs for image classification, and the validation of the results yielded by them. CNN training was conducted

FIGURE 4. Colour variability of four different olive-fruits, of the *Arbequina* variety, collected at the same time within the same orchard.

by transfer learning [19], so pre-trained implementations of the different CNN architectures approached in the present study (which will be later briefly described), available through the MATLAB Deep Learning Toolbox, were used.

The different networks tried were trained by GPU computing, using a single 8GB-NVIDIA GTX 1080 graphic card (Nvidia Corporation, Santa Clara, California, USA).

IV. DEVELOPED METHODOLOGY

A. IMAGE PRE-PROCESSING

As stated before, the main goal of the present study is to provide a computer-vision methodology for the individual classification of olive-fruits, according to their variety, present in images containing multiple individuals, even in touch. A first approach to the problem may consider colour characteristics as a feature to base the classification on. However, for the case of this crop, colour is known to be strongly dependent of the state of maturity of olive-fruits [50], [51]. In this regard, it must be additionally highlighted that, even for any given tree, fruit ripening may vary depending to conditions such as its location within the tree or its degree of exposure to direct sunlight, among other. In order to illustrate this fact, Fig. 4 shows a set of four *Arbequina*-variety olive-fruits, all of them collected in the same orchard, and at the same time, where it can be appreciated the lack of uniformity in terms of colour between the fruits.

As a consequence of these facts, it can be concluded that colour variability induced by fruit maturity may be, at least, comparable to that induced by fruit variety. It is for this reason that colour features were discarded from the beginning, this is from image capture, so morphological characteristics of the fruits were explored as distinctive features for variety categorization.

As detailed in section III-B, fruits were photographed in batches of 50 individuals each. In addition, they were stochastically disposed over the capture area, enabling a scenario in which the fruits may appear touching each other. With the aim of making the methodology able to individually categorise every single olive-fruit present in the images, the first step of pre-processing is to transform initial captures into a suitable set of individual-fruit images.

The methodology proposed by [49] provides an efficient procedure for transforming the starting captures of multiples olive-fruits into binary images, where all the connected components corresponding to the different fruits are accurately separated. This method exploits the high contrast, in terms

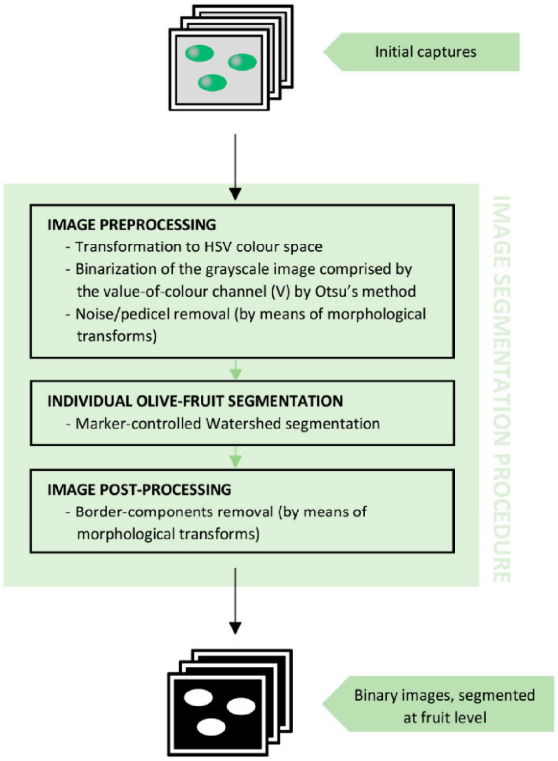


FIGURE 5. Representative diagram of the methodology proposed in [49] for the segmentation of the olive-fruits which appear in an image stochastically positioned.

of luminosity, between the fruits and background to perform the segmentation of the images. This latter, by discriminant analysis-based global thresholding [52]. After that, and by morphological analysis, the resulting binary images are transformed in order to yield a set of markers with which it is carried out an eventual marker-controlled Watershed segmentation [53]. This last step allows to separate those connected components which may appear wrongly fussed, for belonging to fruits that are touching each other in the original capture. Fig. 5 offers a diagram briefly describing this procedure. Fig. 6 shows the result of its application to one of the images initially acquired for this study.

Once images are binarized, and the contained components are correctly isolated by applying this methodology, each of them is extracted and individually included in a new square binary image of 501 pixels high and wide. It should be noted that each component is positioned at the centre, i.e., matching its centre of mass with the centre of the 501×501 frame. Hereafter, the rest of the processing is applied to these individual olive-fruit images.

As the generated images are binary, fruits appear as flat objects. The next step pursues to transform them into olive-fruits with a depth consistent with their shape. To do so, a weighing function based on the distance transform is applied, converting the binary images into a greyscale

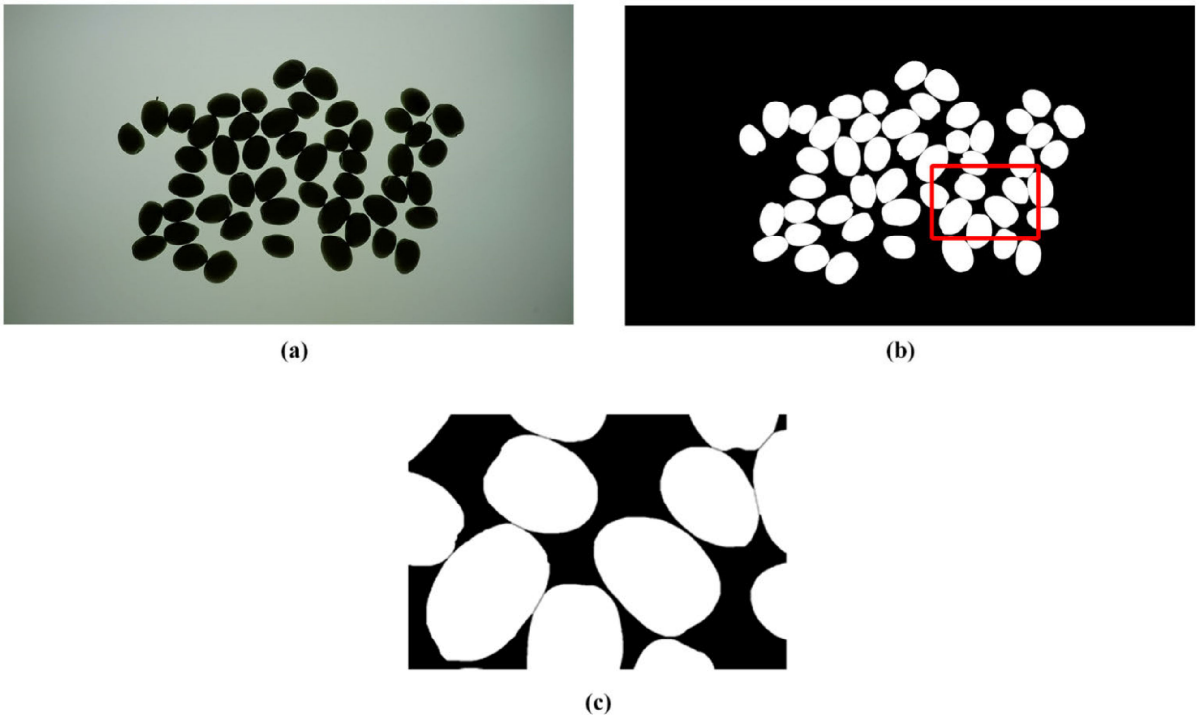


FIGURE 6. (a) Original image; (b) image (a) segmented according with the procedure proposed in [49]; (c) closeup of the red-squared area in image (b).

format. Hence, different grey level values are assigned to each olive-fruit pixel, thus providing the sphericity and three-dimensionality of the fruits [49]. So, being p a foreground pixel of a binary image f of 501×501 pixels in size which contains a single centred olive-fruit, the grey level value v_p assigned to p in the transform image can be computed as follows:

$$v_p = log_3(1 + 2 \times [D(f)](p)), \tag{1}$$

where D refers to the distance function [54] applied to image f, so the expression *[D(f)](p)* can be defined as the Euclidean distance between p and the nearest background pixel, q. Mathematically:

$$[D(f)](p) = min\{d_\varepsilon(p, q) \,|\, f(q) = 0\}, \tag{2}$$

$$d_\varepsilon(p, q) = \sqrt{(p_x - q_x)^2 - (p_y - q_y)^2}. \tag{3}$$

where (p_x, p_y) and (q_x, q_y) are, respectively, the coordinates of the pixels p and q.

The entire process is illustrated in Fig. 7. Note that for the sake of facilitating the visualization of the partial results of the methodology, a sub-image of the original one, which contains a batch of 50 olive-fruits, is used. On the other hand, Fig. 8 shows examples of individual olive-fruit images obtained by the procedure proposed, for each of the varieties considered in the investigation.

B. IMAGE CLASSIFICATION

1) IMAGE DATASET ORGANIZATION

As commented before, the image dataset used throughout the experimentation was generated from photographs of 400 olive-fruits per variety. As proposed in section IV-A, initial captures were processed to obtain individual fruit images, yielding a total of 2,800 of them (400 per each of the seven varieties). 1,050 out of the 2,800 images (150 per variety) were kept for external validation, whereas the resting 1,750 images (250 per variety) were used for training the different CNNs. In order to increase this training dataset, a data augmentation based on image rotation was carried out.

Therefore, per single-fruit image belonging to the training set described hereabove, a rotation transform of 45° was applied. This transformation was repeated by using the resulting image, and so on until seven rotations were performed. Thus, seven extra images were yielded per fruit. Hence, in short, the 1,750 fruits (250 per variety), initially photographed in batches of 50, delivered 14,000 501×501 pixel-resolution images (2,000 per variety), each of them contains a single olive-fruit. On the other hand, it should be noted that the CNN architectures considered were implemented such that the input expected for all of them was 3-channel images. Since the methodology proposed is based on greyscale images, the image-structures obtained by the procedure described hereabove are triplicated and integrated

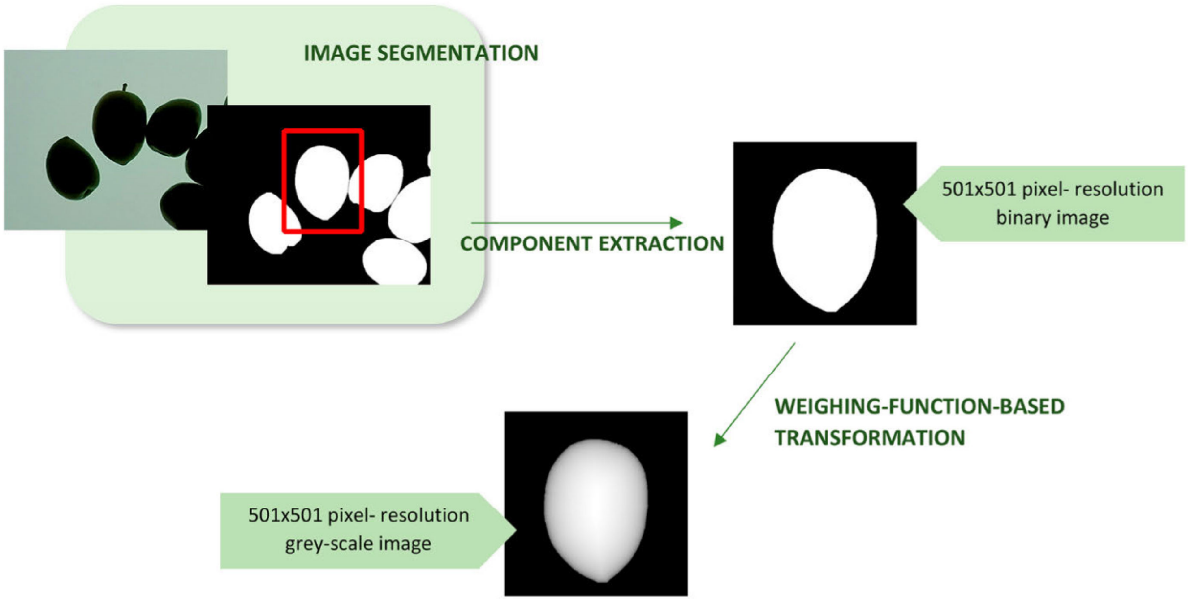


FIGURE 7. Representative diagram of the developed procedure for transforming the initial captures into a set of individual olive-fruit images.

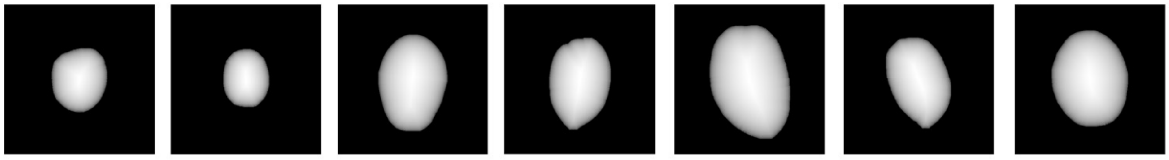


FIGURE 8. Examples of individual olive-fruit images obtained for each of the varieties under study, after transforming initial captures by the proposed procedure: (a) *Arbequina* variety; (b) *Arbosana* variety; (c) *Changlot Real* variety; (d) *Lechín de Sevilla* variety; (e) *Ocal* variety; (f) *Picual* variety; (g) *Verdial de Huévar* variety.

into the same file, so the final images are made up of three channels containing exactly the same information. Fig. 9 illustrates the data augmentation performed.

Table 1 summarizes, per variety, how the corresponding image dataset was organized.

2) CNN TRAINING

Implementations of the considered CNN architectures were trained in order to potentially perform the classification of the olive fruits. To that end, first, the 14,000 individual-fruit training images were appropriately labelled according to their variety. Then, as the CNN architectures were available pre-trained with the thousands of images contained in the ImageNet dataset [37], transfer learning [19] was accomplished to retain convolutional patterns a-priori 'known' by the nets. However, all implementations were modified in order to consider the right number of classification classes. Indeed, ImageNet contains labelled images from 1,000 categories, while the present study considers seven not included in that set, those in accordance with the olive cultivars under study. Therefore, the output layer was conveniently modified in all cases by reducing it to seven classification nodes.

Here below, Table 2 details the configuration parameters and the main training milestones for the different CNN architectures tried.

Each of the CNNs were trained during a different number of epochs, depending on the particularities of each architecture which makes them to converge to an optimum result at a different pace. This convergence was judged by analysing the partial results offered by the loss function evolution during training. In the same way, the mini-batch size used was not necessarily the same for the different architectures either. Indeed, this parameter was adapted to each specific case for optimising the learning time. It should also be noted that during training and before every epoch, the training-data was shuffled. When dividing the training dataset into mini-batches of a same given size, the division was not exact for any case, thus producing residual images not used in the learning

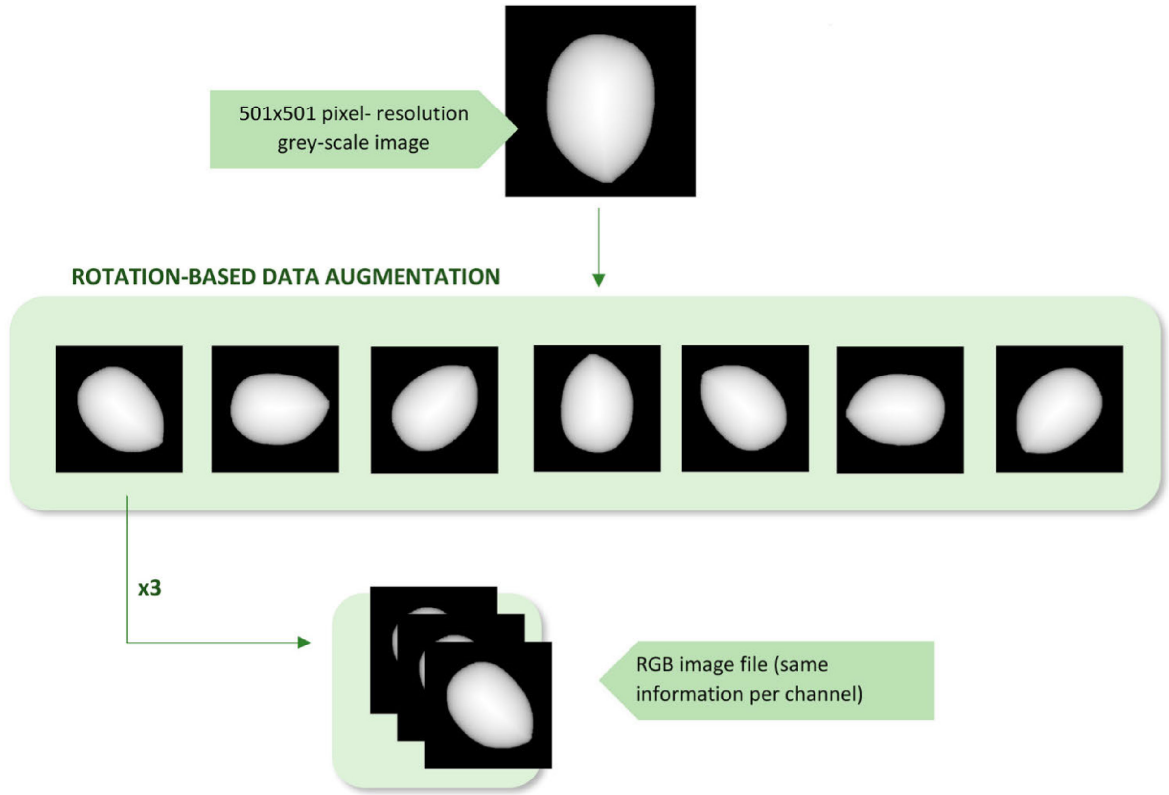


FIGURE 9. Representative diagram of the rotation-based data augmentation performed.

TABLE 1. Materials: organization of the image dataset.

	Number of fruits samples	Initial captures with multiple fruits	Individual-fruit images	Training individual-fruit images	Validation individual-fruit images
i-th variety	400	8[a]	400	2,000[b]	150
Total (7 varieties)	2,800	56[a]	2,800	14,000[b]	1,050

[a] 50 fruits per capture.
[b] After data augmentation.

process. Thanks to data shuffling, the residual images were substituted after every training epoch, thus exploiting the knowledge provided by the whole imagery.

Fig. 10 shows the three maximum filter responses produced by the five convolutional layers of AlexNet, after being trained, when it is excited with sub-image (a). It shows how the outer layers model gross distinctive features of the olive, whereas the inner ones focus on the subtle details. Regarding these last, note how the depth given to the fruit in the pre-processing is captured and exploited by the net for characterisation.

3) CNN VALIDATION

Once the networks were trained, they were tried in classifying the images corresponding to the external validation set. This, in order to assess the goodness of the models yielded after the training process.

The metric proposed for quantifying the performance of the classifiers was based on the ratio between the numbers of fruits correctly categorized within a specific variety, and the total number of them included in the corresponding validation subset. Mathematically:

$$HR_{CN,v} = \frac{h_{CN,v}}{N_v}, \tag{4}$$

where $h_{CN,v}$ is the number of olive-fruits of the variety v correctly categorized by the convolutional neural network CN, and N_v correspond to the total number of fruits (actually images of individual fruits) belonging to that variety, in the external validation set. N_v will be 1,050 in all cases.

TABLE 2. Summary of the configuration of the CNNs, and the main milestones registered from their training.

	AlexNet	InceptionV1	InceptionV3	ResNet-50	ResNet-101	Inception-ResNetV2
Epochs	168	85	67	50	77	74
Iterations	117,648	26,598	37,602	14,090	43,202	103,661
Mini-batch size	20	45	25	50	25	10
Input image size (pixels)	227×227	224×224	299×299	224×224	224×224	299×299
Learning rate	0.001	0.001	0.001	0.001	0.001	0.001

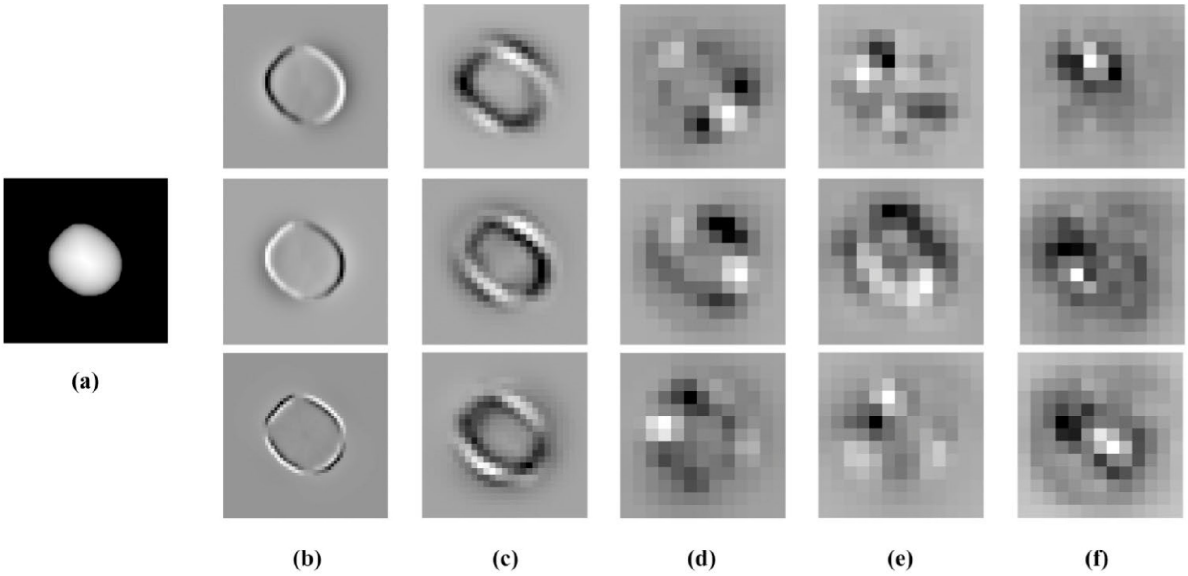


FIGURE 10. Representation of the three maximum filter responses produced by AlexNet to the input image (a); the net was trained with the set of 14,000 individuals kept to this effect. Responses follow the order of their corresponding layers, being (b) those from the outer layer and (f) from the inner. Note how finer details are characterised as the layer is deeper. Also, analyse how the depth given to the fruits in the pre-processing is exploited by the net, especially in the deeper layers, for fruit characterisation.

In addition, the average hit-rate obtained by each classifier, *CN*, is defined as follows:

$$HR_{CN} = \frac{\sum_v HR_{CN,v}}{\#v}, \tag{5}$$

where $\#v$ refers to the number of olive varieties (seven in the present case study).

V. RESULTS AND DISCUSSION

Table 3 shows the results, in terms of the metrics defined in (4) and (5), obtained with the six different CNN architectures. The results were measured from the classifications performed by the different CNNs for the total of 1,050 individual olive-fruit images in the external validation set.

As it can be checked, Inception-ResNetV2 offered the overall best results. Nevertheless, exclusively attending to these outcomes, the differences in terms of accuracy among the different architectures, maybe excepting that given by AlexNet, was not substantial enough to draw definitive conclusions regarding the supremacy of one of them for the case of study. Indeed, in general, performance of all classification models was remarkable, with an average accuracy above 90% in almost all cases.

For a certain image, the CNN output provides a probability value of belonging to each of the seven varieties. The higher the probability given of belonging to the correct class, the better the CNN was able to correctly recognise the variety. Table 4 details the average probability, of belonging to the correct class, and the corresponding standard deviation given by the different architectures when classifying the 150 images of each of the seven varieties; figures are also calculated considering images from all varieties as a whole. It can be identified that Inception-ResnetV2 also outperformed the other alternatives from this perspective; AlexNet was again the less accurate solution.

TABLE 3. CNN implementations tested accuracy, based on results obtained after classifying the images which integrate the external validation set.

Variety	CNN architecture					
	AlexNet ($HR_{CN,v}$)	Inception-ResnetV2 ($HR_{CN,v}$)	InceptionV1 ($HR_{CN,v}$)	InceptionV3 ($HR_{CN,v}$)	Resnet-50 ($HR_{CN,v}$)	Resnet-101 ($HR_{CN,v}$)
Arbequina	0.9333	0.9333	0.9467	0.9533	0.9400	0.9267
Arbosana	0.9133	0.9067	0.9467	0.9400	0.9133	1.0000
Changlot	0.8667	0.9400	0.9200	0.9600	0.9600	0.9333
Lechín	0.8333	0.8867	0.9267	0.9067	0.9200	0.9067
Picual	0.9467	0.8933	0.9600	0.9600	0.9400	0.9867
Ocal	0.9133	0.9333	0.9667	0.9933	0.9667	0.9733
Verdial	0.8867	0.9333	0.9733	0.9600	0.9400	0.9867
Overall	0.8990	0.9181	0.9486	0.9533	0.9400	0.9591

TABLE 4. Average ($\bar{x}$) and standard deviation (σ) of the probability of belonging to the correct class given by the different architectures, measured on the 150 images of each of the seven varieties, and also considering images from all varieties as a whole.

Variety	CNN architecture					
	AlexNet ($\bar{x}, \sigma$)	Inception-ResnetV2 ($\bar{x}, \sigma$)	InceptionV1 ($\bar{x}, \sigma$)	InceptionV3 ($\bar{x}, \sigma$)	Resnet-50 ($\bar{x}, \sigma$)	Resnet-101 ($\bar{x}, \sigma$)
Arbequina	(0.9283, 0.2241)	(0.9333, 0.2199)	(0.9280, 0.2240)	(0.9419, 0.2029)	(0.9477, 0.1910)	(0.9374, 0.2094)
Arbosana	(0.8913, 0.2605)	(0.9894, 0.0573)	(0.9081, 0.2632)	(0.9312, 0.2199)	(0.9342, 0.2007)	(0.9100, 0.2563)
Changlot	(0.8610, 0.3120)	(0.9183, 0.2404)	(0.9294, 0.2283)	(0.9075, 0.2319)	(0.9461, 0.1792)	(0.9561, 0.1583)
Lechín	(0.8270, 0.3448)	(0.9077, 0.2539)	(0.8767, 0.2880)	(0.9076, 0.2286)	(0.8990, 0.2371)	(0.8993, 0.2493)
Picual	(0.9293, 0.228)	(0.9743, 0.1154)	(0.8938, 0.2771)	(0.9465, 0.1762)	(0.9449, 0.1702)	(0.9339, 0.2135)
Ocal	(0.9001, 0.2665)	(0.9757, 0.1346)	(0.9318, 0.2244)	(0.9630, 0.1490)	(0.9701, 0.1005)	(0.9668, 0.1300)
Verdial	(0.8901, 0.2855)	(0.9738, 0.1046)	(0.9324, 0.2184)	(0.9637, 0.1314)	(0.9469, 0.1822)	(0.9639, 0.1409)
Overall	(0.8896, 0.2789)	(0.9532, 0.1783)	(0.9143, 0.2485)	(0.9373, 0.1961)	(0.9413, 0.1852)	(0.9382, 0.2011)

Thus, for the case of the images of *Arbosana*, the average probability of belonging to that variety provided by the net was 0.9894, with a very small standard deviation of 0.0573. This effect can be also found for the case of *Lechín*, *Picual*, *Ocal*, *Verdial* and considering all instances together. As with results from Table 3, differences were not so significant so as to take assertive conclusions. Notwithstanding, both sets of outcomes induce to analogous reflections.

It is worth noting that the CNN based on the AlexNet architecture, closely followed by the InceptionV1-based one, yielded the poorest results in both cases. The rest of CNNs performed better. To note that they all suppose a considerable increase of layers when compared with AlexNet and the first version of the Inception architecture. But nor this fact, neither the use of shortcuts-based architectures, as it happens in some of the cases (note that Resnet-50 yielded better results than Resnet-101), could necessary justify the better accuracy. At least more than the evidence that these four CNNs correspond to more recent and refined neural networks models. Hence, it should be necessary to undertake more research before stating the association between accuracy and the depth of the network. Be that as it may, such a homogeneous and general accurate result registered for all CNNs spotlights that the initial hypothesis of using the morphological modelling of the fruits as distinguishing characteristic is potentially valid. Indeed, given the reasons why colour features might not be appropriate to carry out an accurate variety categorization, the methodology proposed is potentially more suitable in order to achieve a more generalisable solution.

A contingency table is presented in Table 5 for the classifications performed by Inception-ResnetV2. As it can be observed, the net overestimated some classes and underestimated others. Of special interest is the analysis of the actual class *Arbequina*, where six of their images were wrongly

TABLE 5. Contingency table of the classification performed by Inception-ResnetV2 on the set of images for external validation.

Predicted class	Actual class							Predicted instances
	Arbequina	Arbosana	Changlot	Lechín	Picual	Ocal	Verdial	
Arbequina	139 (92.67%)	0 (0.0%)	0 (0.0%)	1 (0.67%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	140
Arbosana	6 (4.0%)	150 (100%)	0 (0.0%)	1 (0.67%)	1 (0.67%)	0 (0.0%)	0 (0.0%)	158
Changlot	1 (0.67%)	0 (0.0%)	140 (93.33%)	2 (1.3%)	0 (0.0%)	1 (0.67%)	1 (0.67%)	145
Lechín	1 (0.67%)	0 (0.0%)	3 (2.0%)	136 (90.67%)	1 (0.67%)	0 (0.0%)	0 (0.0%)	141
Picual	0 (0.0%)	0 (0.0%)	4 (2.67%)	1 (0.67%)	148 (98.67%)	0 (0.0%)	0 (0.0%)	153
Ocal	2 (1.3%)	0 (0.0%)	0 (0.0%)	8 (5.3%)	0 (0.0%)	146 (97.33%)	1 (0.67%)	157
Verdial	1 (0.67%)	0 (0.0%)	3 (2.0%)	1 (0.67%)	0 (0.0%)	3 (2.0%)	148 (98.67%)	156
Actual instances	150	150	150	150	150	150	150	1050

classified as *Arbosana*. This phenomenon could be relevant if it would have also appeared in the contrary case, as it could indicate certain difficulty for the CNN to distinguish between both varieties. However, none of the *Arbosana* images were misclassified as being from *Arbequina*. A similar case occurred when attending to the actual class *Lechín*, as eight of its images were classified as *Ocal*. Notwithstanding, as for the previous case, there were not any images from *Ocal* classified as *Lechín*. This fact strengthen confidence in the capability of the net for discriminating among the different varieties thanks to the image pre-processing applied, and points to enrich the training set with more variability to obtain even better results.

VI. CONCLUSION

The purpose of this study was to assess the viability of a Computer Vision-based methodology to support the automatic and individual classification of olive-fruits, according to the variety they belong to. To that end, it was designed a procedure, based on image processing and analysis and CNNs, for developing a set of image classifiers. To the best to authors' knowledge, the presented algorithm is the first available proposal achieving that end in a non-invasive and accurate manner.

Hence, these image classifiers showed a remarkable behaviour in terms of performance, as high rates of accuracy were obtained in general for all of them, and in particular for the deeper CNN architectures. The outstanding accuracy found for all the different CNN architectures and versions, provides strong signs to validate the initial hypothesis of using morphological modelling of the fruits as distinctive varietal feature, at the expense of discarding colour characteristics.

Despite the quality of the results, further work may be approached by increasing the number of elements the neural networks were trained with. This, to the purpose of testing if it is possible to yield even better accuracy. Along the same lines, future research should include new olive cultivars in order to evaluate the goodness of the methodology proposed, as a general solution to the olive-fruit variety classification.

In addition, it should be underscored that the background the methodology proposed is based on, is aimed to satisfy real-time inspection of olive-fruits when they are transported on a real conveyor belt. By integrating the image processing procedure here presented to transform the initial captures of the fruits, and by applying the classification models computed, an integral solution could be built, potentially applicable to the olive sector, able to perform real time labelling of the olive-fruits, thus making possible their automatic grading and variety-based classification. Particularly, variety-based classification could have an important impact in reducing harvesting costs, as the need of sequentially processing collected fruits of a single variety at a time would disappear; by this way, harvesting could be organised by attending exclusively to the optimisation of operative costs. Additionally, the ability to classify olive-fruits in line could also have an impact in optimising the available space at the warehouse.

ACKNOWLEDGMENT

The authors would like to thank "Virgen de la Oliva" olive-oil cooperative for generously offering their orchards to conduct this work.

REFERENCES

[1] D. Kaniewski, E. Van Campo, T. Boiy, J.-F. Terral, B. Khadari, and G. Besnard, "Primary domestication and early uses of the emblematic olive tree: Palaeobotanical, historical and molecular evidence from the Middle East," *Biol. Rev.*, vol. 87, no. 4, pp. 885–899, Nov. 2012.

[2] P. Vossen, "Olive oil: History, production, and characteristics of the world's classic oils," *Amer. Soc. Horticultural Sci.*, vol. 42, no. 5, pp. 1093–1100, 2007.

[3] Word Table Olive Figures, International Olive Council (IOC). (2019). *World Statistics on Production, Imports, Exports and Consumption*. Accessed: Aug. 6, 2019. [Online]. Available: http://www.internationaloliveoil.org/estaticos/view/132-world-table-olive-figures

[4] Word Olive Oil Figures, International Olive Council (IOC). (2019). *World Statistics on Production, Imports, Exports and Consumption*. Accessed: Aug. 6, 2019. [Online]. Available: http://www.internationaloliveoil.org/estaticos/view/131-world-olive-oil-figures

[5] B. A. Keating, P. S. Carberry, P. S. Bindraban, S. Asseng, H. Meinke, and J. Dixon, "Eco-efficient agriculture: Concepts, challenges, and opportunities," *Crop Sci.*, vol. 50, pp. S-109–S-119, Mar. 2010.

[6] P. C. Struik and T. W. Kuyper, "Sustainable intensification in agriculture: The richer shade of green. A review," *Agronomy Sustain. Develop.*, vol. 37, no. 5, p. 39, Oct. 2017.

[7] H. C. J. Godfray, J. R. Beddington, I. R. Crute, L. Haddad, D. Lawrence, J. F. Muir, J. Pretty, S. Robinson, S. M. Thomas, and C. Toulmin, "Food security: The challenge of feeding 9 billion people," *Science*, vol. 327, no. 5967, pp. 812–818, Feb. 2010.

[8] L. Kitinoja, S. Saran, S. K. Roy, and A. A. Kader, "Postharvest technology for developing countries: Challenges and opportunities in research, outreach and advocacy," *J. Sci. Food Agricult.*, vol. 91, no. 4, pp. 597–603, Mar. 2011.

[9] A. Scheidel and F. Krausmann, "Diet, trade and land use: A socio-ecological analysis of the transformation of the olive oil system," *Land Policy*, vol. 28, no. 1, pp. 47–56, Jan. 2011.

[10] G. Giametta and B. Bernardi, "Olive grove equipment technology. Straddling trees: Mechanized olive harvests," *Adv. Horticultural Sci.*, vol. 24, no. 1, pp. 64–70, 2010.

[11] S. Naik and B. Patel, "Machine vision based fruit classification and grading—A review," *Int. J. Comput. Appl.*, vol. 170, no. 9, pp. 22–34, 2017.

[12] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943.

[13] F. Rosenblatt, "The perceptron, a perceiving and recognizing automaton," Cornell Aeronaut. Lab., Buffalo, NY, USA, Tech. Rep. 85-460-1, 1957.

[14] R. Raina, A. Madhavan, and A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors," in *Proc. 26th Ann. Int. Conf. Mach. Learn.*, Montreal, QC, Canada, Jun. 2009, pp. 873–880.

[15] X.-W. Chen and X. Lin "Big data deep learning: Challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.

[16] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," Apr. 2016, *arXiv:1604.07316*. [Online]. Available: https://arxiv.org/abs/1604.07316

[17] A. Shustanov and P. Yakimov, "CNN design for real-time traffic sign recognition," *Procedia Eng.*, vol. 201, pp. 718–725, Dec. 2017.

[18] D. Chaves, S. Saikia, L. Fernández-Robles, E. Alegre, and M. Trujillo, "Una revisión sistemática de métodos para localizar automáticamente objetos en imágenesc," *Revista Iberoamericana De Automática E Informática Ind.*, vol. 15, no. 3, pp. 231–242, Jun. 2018.

[19] H.-C. Shin, H. R. Roth, M. Gao, L. Le, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[20] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.

[21] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Aug. 2018.

[22] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Comput. Electron. Agricult.*, vol. 153, pp. 69–81, Oct. 2018.

[23] M. Rahnemoonfar and C. Sheppard, "Deep count: Fruit counting based on deep simulated learning," *Sensors*, vol. 17, no. 4, p. 905, 2017.

[24] A. Aquino, B. Millan, M.-P. Diago, and J. Tardaguila, "Automated early yield prediction in vineyards from on-the-go image acquisition," *Comput. Electron. Agricult.*, vol. 144, pp. 26–36, Jan. 2018.

[25] A. Aquino, M. P. Diago, B. Millán, and J. Tardáguila, "A new methodology for estimating the grapevine-berry number per cluster using image analysis," *Biosyst. Eng.*, vol. 156, pp. 80–95, Apr. 2017.

[26] B. Millan, A. Aquino, M. P. Diago, and J. Tardaguila, "Image analysis-based modelling for flower number estimation in grapevine," *J. Sci. Food Agricult.*, vol. 97, no. 3, pp. 784–792, Feb. 2017.

[27] L. A. da Silva, P. O. Bressan, D. N. Gonçalves, D. M. Freitas, B. B. Machado, and W. N. Gonçalves, "Estimating soybean leaf defoliation using convolutional neural networks and synthetic images," *Comput. Electron. Agricult.*, vol. 156, pp. 360–368, Jan. 2019.

[28] T. R. Chavan and A. V. Nandedkar, "AgroAVNET for crops and weeds classification: A step forward in automatic farming," *Comput. Electron. Agricult.*, vol. 154, pp. 361–372, Nov. 2018.

[29] G. Saleem, M. Akhtar, N. Ahmed, and W. S. Qureshi, "Automated analysis of visual leaf shape features for plant classification," *Comput. Electron. Agricult.*, vol. 157, pp. 270–280, Feb. 2019.

[30] Z. Dong, X. Chen, W. Jia, S. Du, K. Muhammad, and S.-H. Wang, "Image based fruit category classification by 13-layer deep convolutional neural network and data augmentation," *Multimedia Tools Appl.*, vol. 78, no. 3, pp. 3613–3632, 2019.

[31] F. J. Rodríguez, A. García, P. J. Pardo, F. Chávez, and R. M. Luque-Baena, "Study and classification of plum varieties using image analysis and deep learning techniques," *Prog. Artif. Intell.*, vol. 7, no. 2, pp. 119–127, 2018.

[32] S. S. Martínez, D. M. Gila, A. Beyaz, J. G. Ortega, and J. G. García, "A computer vision approach based on endocarp features for the identification of olive cultivars," *Comput. Electron. Agricult.*, vol. 154, pp. 341–346, Nov. 2018.

[33] P. Vanloot, D. Bertrand, C. Pinatel, J. Artaud, and N. Dupuy, "Artificial vision and chemometrics analyses of olive stones for varietal identification of five French cultivars," *Comput. Electron. Agric.*, vol. 102, pp. 98–105, Mar. 2014.

[34] A. Beyaz, M. T. özkaya, and D. Içen, "Identification of some spanish olive cultivars using image processing techniques," *Sci. Horticulturae*, vol. 225, pp. 286–292, Nov. 2017.

[35] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.

[36] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, Feb. 2018, Art. no. 7068349.

[37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[38] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," Tech. Rep. 001, Jul. 2012.

[39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[40] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," in *Proc. 4th Int. Conf. Adv. Comput. Commun. Technol.*, Feb. 2014, pp. 5–12.

[41] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[43] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[44] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[46] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. 30th Int. Conf. Mach. Learn.*, Feb. 2013, pp. 1310–1318.

[47] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 6, no. 2, pp. 107–116, 2003.

[48] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," Feb. 2016, *arXiv:1602.07261*. [Online]. Available: https://arxiv.org/abs/1602.07261

[49] J. M. Ponce, A. Aquino, B. Millan, and J. M. Andújar, "Automatic counting and individual size and mass estimation of olive-fruits through computer vision techniques," *IEEE Access*, vol. 7, pp. 59451–59465, 2019.

[50] M. Bouaziz, M. Chamkha, and S. Sayadi, "Comparative study on phenolic content and antioxidant activity during maturation of the olive cultivar Chemlali from tunisia," *J. Agricult. Food Chem.*, vol. 52, no. 17, pp. 5476–5481, 2004.

[51] G. Menz and F. Vriesekoop, "Physical and chemical changes during the maturation of Gordal Sevillana olives (Olea europaea L., cv. Gordal Sevillana)," *J. Agricult. Food Chem.*, vol. 58, pp. 4934–4938, Mar. 2010.

[52] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[53] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. New York, NY, USA: Pearson, 2018.
[54] P. Soille, *Morphological Image Analysis: Principles and Applications*. Berlin, Germany: Springer, 2004.

JUAN M. PONCE was born in Huelva, Andalusia, Spain, in 1982. He received the Engineering degree in computers science from the University of Seville, Seville, in 2009. He is currently pursuing the Ph.D. degree in image analysis with the University of Huelva, Huelva.

After performing different jobs within the software development industry as a Programmer and an Analyst, nowadays he is with the University of Huelva. His research interest includes developing computer-vision-based technology, potentially applicable in food industry and precision agriculture.

ARTURO AQUINO was born in Huelva, Andalusia, Spain, in 1982. He received the Engineering degree in computer science from the University of Seville, in 2006, the M.Sc. degree in computer engineering and networks from the University of Granada, in 2007, and the Ph.D. degree in image analysis from the University of Huelva, in 2011.

Since 2007, he has been with various academic posts at the University of Granada, the University of Huelva, the Research Centre for Mathematical Morphology, Mines ParisTech, and the University of La Rioja. He is currently a Postdoctoral Researcher and a Lecturer with the University of Huelva. Until 2019, he has published more than 40 papers, including articles, book chapters, and contributions to conferences, and he holds two patents. He has directed a doctoral thesis and has participated in eight European, national, and regional research projects. His research interests include image analysis, processing, and understanding.

JOSÉ M. ANDÚJAR was born in Huelva, Spain. He received the Ph.D. degree, in 2000. He is currently a Full Professor of systems engineering and automatic control with the University of Huelva, Spain. Throughout his professional life, he has received 24 awards and academic honors. He has conducted ten doctoral theses with eight prizes, and he holds 12 international patents. He has more than 100 articles published in indexed journals in the ISI Journal Citation Reports. Specifically, he has 51 quartile Q1 publications in 20 different journals; most of these are among the top 10 in their categories, and several are number 1. He has led or co-led more than 50 research projects funded by public institutions and companies. His main research interests include control engineering, renewable energy systems, remote piloted aircraft systems applications, and engineering education.

4.1.4. Congress 1

Estimación automatizada del peso de aceitunas mediante análisis de imagen

J. M. Ponce, A. Aquino, F. Segura, B. Millan, J. M. Andújar

Published in:

SCI 2018 XIV Simposio CEA de Control Inteligente. Málaga, Andalusia, Spain, June 27-29, 2018. PROCEEDINGS. (ISBN: 978-84-09-02635-7)

Editor: Alfonso J. García Cerezo

Reference: pp. 49-54

Year: 2018

XIV Simposio CEA de Control Inteligente. Málaga, 27-29 de Junio 2018

Estimación automatizada del peso de aceitunas mediante análisis de imagen

Juan M. Ponce [a], Arturo Aquino [a], Francisca Segura [a], Borja Millan [a], J. M. Andújar [a]

[a] *Departamento de Ingeniería Electrónica, Sistemas Informáticos y Automática, Universidad de Huelva, Carretera Huelva - Palos de la Frontera s/n, 21819, La Rábida, Palos de la Frontera (Huelva).*

Resumen

El calibrado y selección de productos agrícolas es una actividad de gran relevancia dentro de la industria agroalimentaria. Este estudio, centrado en el sector del olivo, presenta una solución basada en análisis de imagen que permite la estimación automática y no invasiva del peso de un conjunto de aceitunas, a partir de una serie de fotografías de las mismas. Considerando dos variedades distintas de aceituna (Arbequina y Picual), se ha desarrollado un algoritmo de segmentación a partir del cual se extrae la información necesaria para computar modelos de predicción de peso. Una vez aplicados dichos modelos sobre los correspondientes conjuntos de validación, se ha podido comprobar, a través del cálculo de la raíz del error cuadrático medio (RMSE) cometido, la eficacia del método propuesto y su validez como base para el desarrollo de un sistema de calibrado de aceitunas de bajo coste basado en visión artificial.

Palabras Clave:
aceituna, peso, segmentación, visión artificial, estimación.

Automated Olive-fruit weight estimation by means of image-analysis.

Abstract

The sizing and sorting of agricultural commodities is a high relevance activity in food industry. This study, focused on the olive farming sector, presents a solution based on image analysis which allows the automatic and non-invasive estimation of the weight of a set of olive fruits. Considering two different varieties of olive fruits (Arbequina and Picual), a segmentation algorithm able to extract from images the needed information to compute the weight prediction models was developed. The effectiveness of the proposed method was assessed by calculating the root-mean-square error (RMSE) produced by the models when applied to the corresponding external validation sets. The measured results show evidences of viability as a base to the development of a low-cost olive fruit grading system based on machine vision.

Keywords:
olive fruit, weight, segmentation, machine vision, estimation.

1. Introducción

El cultivo del olivo se ha convertido en una actividad agrícola de gran peso y relevancia en la actualidad. Con una gran presencia y tradición en la cuenca mediterránea, zona que ha concentrado históricamente a los principales países productores, su laboreo se ha extendido alrededor de todo el mundo en las últimas décadas, y ya es practicada en lugares tan dispares como el continente americano, Australia y Nueva Zelanda, e incluso en China o el sudeste asiático (IOC, 2018). La principal razón de dicha expansión pasa por el espectacular auge en popularidad y el creciente consumo que, en los últimos años, han experimentado los principales productos que derivan de la industria sustentada por dicho cultivo: las aceitunas de mesa y el aceite de oliva. No en vano, y de acuerdo con datos del Consejo Oleícola Internacional, el consumo de aceitunas de mesa se ha visto incrementado en más de un 170% en los últimos 25 años, y las expectativas de consumo de aceite de oliva para el curso 2016/17 hablaban de 2.800.000 toneladas a nivel mundial (IOC, 2018).

Con el objeto de satisfacer esta creciente demanda, toda la industria generada alrededor del olivo ha de afrontar

ISBN 978-84-09-02635-7

numerosos retos, muchos de ellos de carácter tecnológico, ante una imperativa necesidad de modernización de todos sus procesos.

El tamaño y peso, además de otras características físicas, de los productos derivados de la actividad agrícola son datos de gran valor para su industria. La clasificación y/o selección de los frutos en base a estos parámetros es tarea de especial importancia en según qué cultivos, siendo de vital relevancia en el caso del fruto del olivo. Así, en el caso de, por ejemplo, la aceituna de mesa, el proceso de calibrado, y la precisión del mismo, tienen un especial impacto en el producto final, dadas las expectativas de homogeneidad en cuanto a tamaño que se le presupone al consumidor final.

Frente a las tradicionales soluciones de carácter mecánico, los sistemas de visión artificial comienzan a ganar interés dentro de la industria de transformación agroalimentaria (Benalia *et al.*, 2015; Gómez-Sanchís *et al.*, 2013; Pallottino *et al.*, 2013a, b). Ante este escenario, este tipo de tecnología puede ser de gran valía dentro del sector oleícola ya que no sólo abre la puerta a sistemas de clasificación y selección del fruto menos agresivos con éste, sino que además hacen posible la adquisición de una información antes inaccesible.

El presente estudio propone un sistema de estimación de peso basado en análisis de imagen. En concreto, el algoritmo desarrollado es capaz de estimar el peso individual de las aceitunas contenidas en una imagen digital, empleando técnicas basadas en morfología matemática y segmentación por umbralización.

2. Materiales y metodología

2.1. Adquisición de imágenes y datos de referencia

Para el presente trabajo se cosecharon, en enero de 2018, muestras de aceituna de las variedades Arbequina y Picual. De ellas se seleccionaron 200 individuos por variedad para generar los conjuntos de estudio. En el caso de la variedad Arbequina, se establecieron los siguientes 5 conjuntos: A1, A2 y A3, con 40 individuos por conjunto, A4 con 50 individuos, y A5 con 30. Para la variedad picual, se formaron 4 conjuntos, P1, P2, P3 y P4, con 50 individuos cada uno.

Establecidos los conjuntos de estudio, se procedió al fotografiado de los mismos. Para ello, sus frutos integrantes fueron dispuestos de forma ordenada, y separados los unos de los otros, sobre un tapete blanco (véase un ejemplo de captura en la Fig. 1). Las imágenes fueron tomadas con la cámara digital sin espejo LUMIX DMC-GH4 (Panasonic, Kadoma, Osaka, Japón). Ésta fue configurada en modo manual, con una apertura de f/8, un tiempo de exposición de 1/500s, una sensibilidad ISO de 400 y una distancia focal de 14mm. Para iluminar la escena, se utilizaron dos lámparas halógenas de 500W, con una temperatura de color de 3300k. La cámara fue situada por encima del tapete sobre el que se situaron las aceitunas a fotografiar mirando perpendicularmente a éste. Los focos se situaron a la misma altura, orientados hacia el punto de enfoque de la cámara. Las capturas fueron almacenadas en formato JPG, con una resolución de 4608×2592 píxeles, una densidad de 180 píxeles por pulgada y una profundidad de color de 24bits.

Figura 1: Ejemplo de imagen capturada para uno de los conjuntos de aceituna bajo estudio.

Una vez finalizada la captura de las imágenes de los conjuntos, se realizó el pesado individualizado (en gramos, g) de cada fruto usando la balanza de precisión KERN PCB 3500-2 (KERN & Sohn GmbH, Balingen, Alemania). Finalmente, cada medición se asoció unívocamente, mediante etiquetado, a su correspondiente aceituna fotografiada.

2.2. Algoritmo de análisis de imagen

La metodología propuesta se basa en la extracción automática de características de las imágenes que son descriptivas del peso de las aceitunas. Para ello, el algoritmo desarrollado emplea técnicas basadas en morfología matemática y segmentación por umbralización.

2.2.1. Preprocesamiento

En primer lugar, las imágenes se escalan a un 40% de su tamaño original mediante interpolación bicúbica, ello para reducir la carga computacional de los algoritmos. Seguidamente, se aplica un filtrado gaussiano para la atenuación de ruido de sal y pimienta, con kernel de 5×5 y desviación típica de 0.8. Finalmente, las imágenes se transforman, desde su espacio de color nativo RGB, a una representación en base al espacio de color HSV (del inglés *Hue*, *Saturation*, *Value*). Esta elección estuvo motivada por las características propias de las imágenes capturadas en laboratorio, así como por la propia naturaleza del experimento. Y es que, en efecto, abordar una segmentación basada en características de color carecía de sentido aquí, al no existir garantías de homogeneidad entre las distintas aceitunas en cuanto a matiz (hecho especialmente destacable en la variedad Arbequina). Sin embargo, el utilizar un tapete de color blanco como fondo de escena propició que sus píxeles presentaran *valores de color* o *luminosidad* considerablemente mayores a los de los píxeles pertenecientes a las aceitunas. Por todo ello, se decidió la conversión de las imágenes al espacio HSV, que proporciona la información de luminosidad, en el canal *V*, desacoplada de otras características. Cabe mencionar que hay más espacios de color potencialmente válidos para este trabajo por ofrecer la misma característica, como por ejemplo los CIELAB (canal *L*) (Connolly and Fliess, 1997) o HSL (canal *L*) (Sonka *et al.*, 2014), entre otros.

2.2.2. Segmentación de las aceitunas

Una vez ejecutada la transformación al espacio HSV, el canal correspondiente a la información del valor de color (*value* en inglés) es almacenada en una matriz independiente, y tratada a partir de este punto como una imagen en escala de grises que, tras diversas transformaciones, será finalmente la imagen binarizada.

ISBN 978-84-09-02635-7

XIV Simposio CEA de Control Inteligente. Málaga, 27-29 de Junio 2018

Para la matriz contenedora de dicha componente, se realiza una inversión de los valores de sus elementos en base al máximo valor *de gris* contemplado, esto es 255. Así, sea *V* la imagen definida en el intervalo [0,…,255] (8 bits) correspondiente al canal *Value* del espacio de color HSV, la imagen V_{INV} es la que resulta de la siguiente operación:

$$V_{INV} = 255 - V$$

Con esto, el objetivo perseguido es que los valores de gris mayores (más próximos al blanco) sean los propios de los píxeles que se corresponden con las aceitunas, y que, por tanto, el fondo de la imagen pase a ser la parte más oscura de ésta.

A continuación, para obtener una estimación de fondo, se realiza una apertura morfológica sobre la matriz resultante de la inversión de valores anterior, V_{INV}:

$$I_\gamma = \gamma_\beta(I_{INV}) = \delta_\beta(\varepsilon_\beta(I_{INV}))$$

donde β es un elemento estructurante circular de radio 50 píxeles (lo suficientemente grande como para contener completamente una aceituna de la imagen), y δ y ε las operaciones morfológicas básicas de dilatación y erosión, respectivamente (Soille, 2013).

En búsqueda de una imagen de alto contraste entre las aceitunas y el fondo, la estimación de fondo calculada es sustraída a la matriz de la que origina:

$$I_C = I_{inv} - I_\gamma$$

donde I_C es la imagen de *alto contraste*.

Como última transformación previa a la segmentación definitiva, se aplica una operación de relleno por difusión para rellenar los "huecos" que aparecen sobre las aceitunas, y que se corresponde con los puntos de máximo brillo, consecuencia del fenómeno de reflectividad máxima de la luz sobre la superficie convexa de los frutos. Complementariamente, y debido al procesamiento previo aplicado, este operador promueve también la homogeneización de los valores de gris del fondo de la imagen. Matemáticamente:

$$I_F = R^\varepsilon_{I_C}(I_m),$$

donde R es la operación de reconstrucción morfológica, consistente en la erosión (ε) iterativa de la imagen I_m con respecto a I_C, con un elemento estructurante unitario, hasta la idempotencia (Soille, 2013):

$$R^\varepsilon_{I_C}(I_m) = \varepsilon^i_{I_C}(I_m), \text{ siendo}$$

$$I_m(x,y) = \begin{cases} I_C(x,y) & si\ (x,y) \in borde\ de\ I_C \\ \max(I_C) & en\ otros\ casos \end{cases}, \text{ y}$$

$$\varepsilon^i_{I_C}(I_m) = \varepsilon^{i+1}_{I_C}(I_m), \qquad \varepsilon^1_{I_C}(I_m) = \varepsilon_{\beta=1}(I_m) \vee I_C$$

Finalmente, la imagen resultante I_F, aún en escala de grises, es sometida al proceso de binarización por umbralización. Para ello, se establece como umbral su mínimo valor de gris, el cual representa a los píxeles pertenecientes al fondo (como ya se ha comentado, el propósito de algunas de las transformaciones realizadas es el de conseguir la uniformidad en cuanto a valor para dichos píxeles). Así, los valores por encima de dicho mínimo se asocian con píxeles de aceituna y se segmentan del fondo (ver Figura 2):

$$I_{BIN}(x,y) = \begin{cases} 255 & si\ I_F(x,y) > \min(I_F) \\ 0 & en\ otros\ casos \end{cases}$$

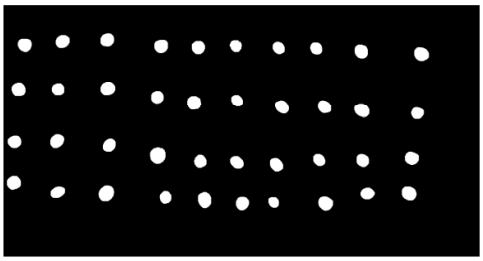


Figura 2: Resultado de la segmentación de la imagen mostrada en la Fig. 1.

2.2.3. Postprocesamiento

Como último paso, se aplica una apertura morfológica con un elemento estructurante pequeño para "limpiar" la imagen binaria resultante de aquellas componentes conexas con un área anormalmente pequeña como para corresponder a una aceituna segmentada:

$$I_{DEF} = \gamma_\beta(I_{BIN}),$$

donde β es un elemento estructurante circular de 10 píxeles de radio.

2.2.4. Evaluación de la calidad de la segmentación

Se realizó una segmentación a mano para cada una de las capturas, utilizando un editor de gráficos, con el objetivo de obtener un resultado lo más cercano posible a una segmentación ideal (*Gold Standard*). El conjunto de imágenes de referencia resultante fue utilizado para, mediante comparación con las generadas por el algoritmo, determinar la calidad de la segmentación proporcionada por éste. Mediante dicha comparación, se hace posible el computo de las métricas *RC* (*Recall*) y *PC* (*Precision*) propuestas a continuación:

$$RC = \frac{TP}{TP + FN} \qquad PC = \frac{TP}{TP + FP}$$

Dónde *TP* (*True Positives*), se refiere al número de píxeles categorizados por el algoritmo como pertenecientes a una aceituna (son píxeles blancos en la imagen binarizada) que también lo son según la segmentación ideal (*Gold Standard*). *FN* (*False Negatives*), se corresponde con el número de píxeles erróneamente clasificados como no pertenecientes a alguna aceituna. De la misma forma, FP (*False Positives*), se refiere al número de píxeles considerados erróneamente como pertenecientes a alguno de los frutos, en base a la clasificación propuesta por la segmentación *Gold Standard*.

ISBN 978-84-09-02635-7

XIV Simposio CEA de Control Inteligente. Málaga, 27-29 de Junio 2018

Junto a *RC* y *PR*, se propone una tercera métrica, $F_{measure}$, basada en los mismos principios, y que evalúa la calidad global de la segmentación:

$$F = 2\,\frac{PC\ x\ RC}{PC + RC}$$

2.3. Extracción de información y desarrollo del modelo de estimación

Una vez generadas las distintas segmentaciones a partir de las capturas iniciales, es necesario extraer de éstas la información necesaria para realizar un estudio de correlación con las mediciones de peso tomadas previamente en laboratorio. Dentro de cada imagen binaria, para cada región o componente conexa correspondiente a un fruto, se calcula el número de píxeles que la integran; éste es el valor a enfrentar con el peso medido en gramos de la aceituna correspondiente.

Una vez es adquirido este conjunto de datos, para cada una de las dos variedades consideradas en el experimento, se procede como sigue: del conjunto de 200 frutos que comprende la muestra de estudio, se toman 50, representando la variabilidad del conjunto de la muestra, para conformar un conjunto de entrenamiento. Mediante un análisis de regresión, basado en la relación de dependencia entre los pesos de las aceitunas de dicho conjunto y el número de píxeles que corresponden a las mismas (según lo aportado por la segmentación), se obtiene la recta de regresión que determina, mediante la ecuación que la describe, el modelo predictivo obtenido. Este modelo es aplicado sobre los 150 frutos restantes, actuando dicho subgrupo como conjunto de validación externo, a partir del cual se evalúa la calidad del modelo.

Insistiendo en que este procedimiento es realizado de forma independiente para cada una de las variedades, en última instancia, se contempla la generación de un único modelo válido para ambas de forma conjunta. Para ello, se hace uso de los dos conjuntos de entrenamiento establecidos anteriormente, abordando el análisis de regresión a partir de la unión de ambos, y aplicando el resultado sobre la suma de los conjuntos de validación externa correspondientes.

2.3.1. Evaluación de la calidad de los modelos de estimación

Una vez calculados los modelos de predicción, y aplicados éstos sobre las poblaciones correspondientes, se llevó a cabo una evaluación final de las estimaciones realizadas. Para ello se recurrió al cálculo de la raíz del error cuadrático medio, según lo establecido a continuación:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

Donde $\hat{y}_i$ y y_i son, respectivamente, el valor predicho y el valor real del peso, para el fruto *i-ésimo* de la muestra tratada, de un total de *n* elementos.

También se realizó el cálculo del error cuadrático relativo a la media como porcentaje, en base a la siguiente definición:

$$SE = \frac{RMSE}{\frac{\sum_{i=1}^{n} y_i}{n}}\ x\ 100$$

Finalmente, se determinó el error relativo (como porcentaje), una vez más para cada modelo y la población de aceitunas correspondiente. Dicho error se calculó según se indica a continuación:

$$|E| = \frac{|\sum_{i=1}^{n}(\hat{y}_i - y_i)|}{\sum_{i=1}^{n} y_i}\ x\ 100$$

3. Resultados y discusión

3.1. Resultados de la evaluación de calidad de la segmentación

Como se ha comentado anteriormente, la viabilidad del algoritmo para la segmentación ha sido testada en base a la comparación de las imágenes generadas por éste con segmentaciones *presumiblemente* ideales (*Gold Standard*), generadas a mano. Esto, a través de las métricas propuestas en el apartado anterior. En la Tabla 1 se muestran los resultados obtenidos, mostrando los valores para cada una de las segmentaciones realizadas. Según los datos recogidos en ella, se puede concluir la solidez del algoritmo de segmentación propuesto. Destacar que no se observan importantes diferencias entre las distintas variedades, manteniéndose los resultados obtenidos siempre dentro de unos márgenes similares. Esta uniformidad respalda la decisión de no basar el proceso de segmentación en características de color (dadas las significativas diferencias en cuanto a tono entre las dos variedades bajo estudio), y sugiere su correcto funcionamiento tratando nuevas variedades, o frutos de ésta en distintos estados de maduración. Otra de las bondades aportadas por el método es su eficacia a la hora de detectar únicamente aquellas componentes conexas que se corresponden con aceitunas, descartando en la segmentación final cualquier componente anómala que no pueda corresponderse con algún fruto. Esto hace posible determinar de forma exacta el número de frutos que aparecen en cada imagen.

Tabla 1: Resultados de la evaluación de la calidad y precisión de las segmentaciones realizadas por el algoritmo propuesto.

muestra	*RC*	*PC*	$F_{measure}$
A1	1	0,8633652	0,9266731
A2	1	0,87684168	0,93438002
A3	1	0,88152872	0,93703456
A4	1	0,89458439	0,9443595
A5	1	0,8669487	0,92873329
promedio	1	0,87665374	0,93423609
P1	0,99997252	0,87503607	0,93334188
P2	1	0,89381748	0,94393202
P3	0,99999426	0,88897784	0,94122379
P4	1	0,90662967	0,9510286
promedio	0,99999336	0,89111527	0,94238157

3.1. Resultados de la evaluación de calidad de los modelos de predicción

La correlación existente entre los resultados de segmentación (las áreas como número de pixeles de las componentes conexas que pertenecen a los frutos) y las

ISBN 978-84-09-02635-7

mediciones de peso de las aceitunas correspondientes, puede observarse en los gráficos mostrados en las figuras 2, 3 y 4. Nótese los altos valores de correlación obtenidos en la obtención de los modelos de estimación, en todos los casos, por encima de 0.9.

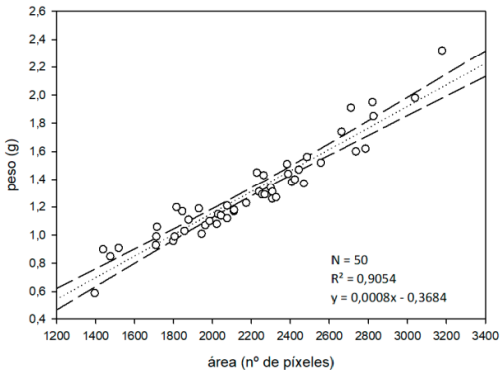


Figura 2: Obtención del modelo de predicción de peso para la variedad Arbequina a partir del conjunto de entrenamiento.

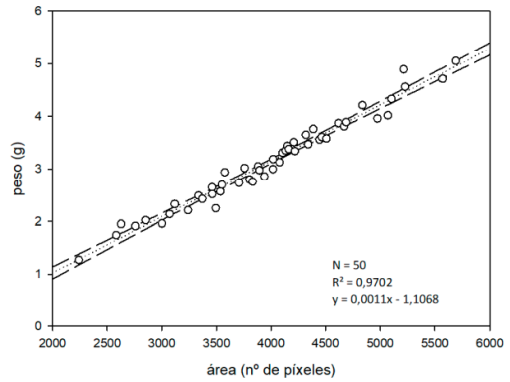


Figura 3: Obtención del modelo de predicción de peso para la variedad Picual a partir del conjunto de entrenamiento.

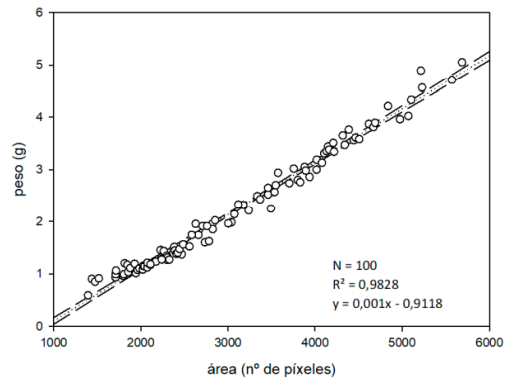


Figura 4: Obtención del modelo de predicción de peso independiente de la variedad a partir del conjunto de entrenamiento.

Por otra parte, con el objeto de ilustrar la calidad predictiva de los modelos, se muestran en la Tabla 2 los resultados de su aplicación sobre los conjuntos de validación externa.

Tabla 2: Resultados de las distintas mediciones computadas para evaluar la calidad de los modelos de estimación, tras ser estos aplicados sobre los conjuntos de validación externa correspondientes (según variedad).

Variedad	*RMSE (g)*	*SE (%)*	*|E| (%)*
Arbequina	0,131495	10,36649	7,406528
Picual	0,256875	8,105015	5,602251
Ambas variedades (modelo global)	0,154839	6,97819	1,105803

Estos resultados ponen de relieve la viabilidad del método, esto es, la estimación del peso de distintas aceitunas capturadas en una imagen a partir del número de píxeles que les corresponden en ésta, y por extensión, el uso del análisis de imagen para dar solución al problema planteado.

Atendiendo a que la calidad de la segmentación para ambas variedades es análoga, las leves diferencias en los valores de error cometido en las estimaciones se pueden explicar por las propias diferencias morfológicas existentes entre los frutos de estas variedades. El método aproxima el problema mediante la proyección de la aceituna en un plano, y considera que cada píxel tiene una aportación equitativa en su peso. Esta aproximación, aunque válida a tenor de los resultados obtenidos, puede empobrecer éstos ligeramente para según que variedades, en base a determinadas características morfológicas, tales como su esfericidad.

Por otro lado, es interesante destacar la ausencia de indicios de un beneficio claro en el uso de modelos específicos para cada variedad en detrimento de un modelo general. Ello potencia la usabilidad y generalización de la metodología desarrollada.

Finalmente, cabe también reseñar que no hay evidencias claras que indiquen la necesidad de explorar modelos no lineales más complejos y de más difícil manejo.

4. Conclusiones

En el presente estudio, se ha desarrollado un método para la estimación del peso de aceitunas mediante análisis de imagen. Los resultados obtenidos muestran la solidez y precisión de éste, y respaldan el uso de este tipo de técnicas en la implementación de sistemas de calibrado y selección, basados en visión artificial, dentro la industria olivícola.

Cabe destacar la estabilidad en el comportamiento del método al tratar con dos tipos distintos de aceitunas y con distintos estados de maduración, hecho que pone de relieve sus posibilidades como estimador independiente de la variedad. Por otra parte, en todas las imágenes testeadas, el número de componentes conexas segmentadas por el algoritmo coincidió exactamente con el número de frutos. Esta eficacia apoya la confianza en el uso de la solución presentada también para realizar un control preciso del número de aceitunas procesadas en un hipotético sistema real basado en ella.

El presente trabajo admite cierto margen para ser extendido en investigaciones futuras. Así, cabe la posibilidad de no sólo

ISBN 978-84-09-02635-7

acometer la estimación de los valores de peso de los frutos, sino también el calibre de los mismos. Por otra parte, resultaría interesante cuantificar el impacto de la ponderación de la contribución de los píxeles de aceituna en la estimación de su peso. Por último, es necesario corroborar los resultados obtenidos sobre un conjunto de datos más extenso, y considerando un mayor número de variedades de aceituna, al objeto verificar su generalidad.

Agradecimientos

Este trabajo se ha realizado en el contexto del proyecto TecnOlivo, aprobado en la primera convocatoria del Programa Interreg V-A España - Portugal (POCTEP) 2014-2020, y cofinanciado con fondos FEDER.

Referencias

Benalia, S., Bernardi, B., Cubero, S., Leuzzi, A., Larizza, M., Blasco, J., 2015. Preliminary trials on hyperspectral imaging implementation to detect mycotoxins in dried figs. Chem. Eng. Trans. 44, 157–162.

Connolly, C., Fliess, T. A, 1997. Study of efficiency and accuracy in the transformation from RGB to CIELAB color space. IEEE Trans. Image Process. 6, 1046–1047.

Gómez-Sanchis, J., Blasco, J., Soria-Olivas, E., Lorente, D., Escandell-Montero, P., Martínez-Martínez, J.M., Aleixos, N., 2013. Hyperspectral LCTF-based system for classification of decay in mandarins caused by Penicillium digitatum and Penicillium italicum using the most relevant bands and non-linear classifiers. Postharvest Biol. Technol. 82, 76–86.

International Olive Oil Council (IOC), 2018. IOC updates series of world statistics on production, imports, exports and consumption. Available: http://www.internationaloliveoil.org/estaticos/view/131-world-olive-oil-figures.

Pallottino, F., Costa, C., Antonucci, F., Menesatti, P., 2013a. Sweet cherry freshness evaluation through colourimetric and morphometric stem analysis: two refrigeration systems compared. Acta Aliment. 42, 428–436.

Pallottino, F., Menesatti, P., Lanza, M.C., Strano, M.C., Antonucci, F., Moresi, M., 2013b. Assessment of quality-assured Tarocco orange fruit sorting rules by combined physicochemical and sensory testing. J. Sci. Food Agric. 93 (1176), 1183.

Soille, P. (2013). Morphological image analysis: principles and applications. Springer Science & Business Media, Germany.

Sonka, M., Hlavac, V., Boyle, R, 2014. Image processing, analysis, and machine vision. Cengage Learning, USA.

ISBN 978-84-09-02635-7

4.1.5. Congress 2

ESTIMACIÓN AUTOMATIZADA DEL PESO Y CALIBRE DE ACEITUNAS MEDIANTE ANÁLISIS DE IMAGEN

J. M. Ponce, A. Aquino, F. Segura, B. Millan, J. M. Andújar

Published in:

XXXIX JORNADAS DE AUTOMÁTICA. Badajoz, Extremadura, Spain, September 5-7, 2018. PROCEEDINGS. (ISBN: 978-84-09-04460-3)

Editor: Inés Tejado Balsera [et al.]

Reference: pp. 958-966

Year: 2018

Actas de las XXXIX Jornadas de Automática, Badajoz, 5-7 de Septiembre de 2018

ESTIMACIÓN AUTOMATIZADA DEL PESO Y CALIBRE DE ACEITUNAS MEDIANTE ANÁLISIS DE IMAGEN

Juan M. Ponce
[1]Departamento de Ingeniería Electrónica, Sistemas Informáticos y Automática, Universidad de Huelva,
Carretera Huelva - Palos de la Frontera s/n,
21819, La Rábida, Palos de la Frontera (Huelva), jmponce.real@diesia.uhu.es

Arturo Aquino[1], arturo.aquino@diesia.uhu.es; Francisca Segura[1], francisca.segura@diesia.uhu.es; Borja Millán[1], borja.millan@diesia.uhu.es; J. M. Andújar[1], andujar@diesia.uhu.es

Resumen

El calibrado y selección de productos agrícolas es una actividad de gran relevancia dentro de la industria agroalimentaria. Este estudio, centrado en el sector del olivo, presenta una solución basada en análisis de imagen que permite la estimación automática y no invasiva del peso y calibre (ejes de simetría mayor y menor) de un conjunto de aceitunas, a partir de una serie de fotografías de las mismas. Utilizando dos variedades distintas de aceituna (Arbequina y Picual), se ha desarrollado un algoritmo de segmentación, a partir del cual se extrae la información necesaria para computar modelos de estimación para cada uno de los parámetros considerados. Una vez aplicados dichos modelos sobre los correspondientes conjuntos de validación, se ha podido comprobar, a través del cálculo de la raíz del error cuadrático medio (RMSE) cometido, la eficacia del método propuesto y su validez como base para el desarrollo de un sistema de calibrado de aceitunas de bajo coste basado en visión artificial.

Palabras clave: aceituna, peso, calibre, segmentación, visión artificial, estimación.

1 INTRODUCCIÓN

El cultivo del olivo se ha convertido en una actividad agrícola de gran peso y relevancia en la actualidad. Con gran presencia y tradición en la cuenca mediterránea, zona que ha concentrado históricamente a los principales países productores, su laboreo se ha extendido alrededor de todo el mundo en las últimas décadas, y ya es practicada en lugares tan dispares como el continente americano, Australia y Nueva Zelanda, e incluso en China o el sudeste asiático [4]. La principal razón de dicha expansión pasa por el espectacular auge en popularidad y el creciente consumo de los principales productos que derivan de la industria sustentada por dicho cultivo: las aceitunas de mesa y el aceite de oliva. No en vano, y de acuerdo con datos del Consejo Oleícola Internacional (COI), el consumo de aceitunas de mesa se ha visto incrementado en más de un 170% en los últimos 25 años [4], y las expectativas de consumo de aceite de oliva, para el curso 2017/18, son de 3.000.000 toneladas a nivel mundial [12]. Con el objeto de satisfacer esta creciente demanda, toda la industria generada alrededor del olivo ha de afrontar numerosos retos, muchos de ellos de carácter tecnológico, ante la imperativa necesidad de modernización de todos sus procesos.

El tamaño y peso, además de otras características físicas, de los productos derivados de la actividad agrícola son datos de gran valor para su industria [6]. La clasificación y/o selección de los frutos en base a estos parámetros es tarea de especial importancia en según qué cultivos, siendo de vital relevancia en el caso del fruto del olivo [5, 9]. Así, en el caso de, por ejemplo, la aceituna de mesa, el proceso de calibrado, y la precisión del mismo, tienen un especial impacto en el producto final, dadas las expectativas de homogeneidad en cuanto a tamaño que se le presupone al consumidor final.

Frente a las tradicionales soluciones de carácter mecánico, los sistemas de visión artificial comienzan a ganar interés dentro de la industria de transformación agroalimentaria [1,3,7,8]. Ante este escenario, este tipo de tecnología puede ser de gran valía dentro del sector oleícola ya que no sólo abre la puerta a sistemas de clasificación y selección del fruto menos agresivos con éste, sino que además hacen posible la adquisición de una información antes inaccesible.

El presente estudio propone un sistema de estimación de peso y calibre (ejes mayor y menor) basado en análisis de imagen. En concreto, el método presentado es capaz de estimar el peso, así como los diámetros mayor y menor de cada una de las aceitunas contenidas en una imagen digital, empleando técnicas

basadas en morfología matemática y segmentación por umbralización. Dicha metodología es desarrollada a lo largo de la siguiente sección. Así, en primer lugar, se recoge en detalle cómo se ha realizado la captura de las imágenes, además de las mediciones de peso y calibre de cada uno de los frutos. Seguidamente, se exponen los distintos pasos acometidos por el algoritmo de segmentación propuesto para la transformación y binarización de dichas imágenes, y se establecen los criterios para evaluar la calidad de las segmentaciones obtenidas. Finalmente, es expuesto el proceso de extracción de datos característicos del peso y calibre desde las imágenes segmentadas, y la generación de los modelos de estimación a partir de éstos y de las mediciones previamente realizadas. En la Sección 3 son presentados y discutidos los resultados obtenidos, tanto para las segmentaciones realizadas como para los modelos de estimación computados. En el último apartado (Sección 4), a modo de conclusión, se destacan algunas implicaciones relacionados con los resultados obtenidos, y se sugieren algunos aspectos a tener en consideración en futuras investigaciones.

2 MATERIALES Y METODOLOGÍA

2.1 ADQUISICIÓN DE IMÁGENES Y DATOS DE REFERENCIA

Para el presente trabajo se cosecharon, en enero de 2018, muestras de aceituna de las variedades Arbequina y Picual. De ellas se seleccionaron 200 individuos por variedad para formar los conjuntos de estudio. En el caso de la variedad Arbequina, se establecieron los siguientes 5 conjuntos: A1, A2 y A3, con 40 individuos por conjunto, A4 con 50 individuos, y A5 con 30. Para la variedad picual, se formaron 4 conjuntos, P1, P2, P3 y P4, con 50 individuos cada uno.

Establecidos los conjuntos de estudio, se procedió al fotografiado de los mismos. Para ello, sus frutos integrantes fueron dispuestos de forma ordenada, y separados los unos de los otros, sobre un tapete blanco (véase un ejemplo de captura en la Fig. 1). Las imágenes fueron capturadas con una cámara digital sin espejo, y equipada con sensor NMOS, modelo LUMIX DMC-GH4 (Panasonic, Kadoma, Osaka, Japón). Ésta fue configurada en modo manual, con una apertura de f/8, un tiempo de exposición de 1/500s, una sensibilidad ISO de 400 y una distancia focal de 14mm. Para iluminar la escena, se utilizaron dos lámparas halógenas de 500W, con una temperatura de color de 3300k. La cámara fue montada por encima del tapete sobre el que se situaron las aceitunas a fotografiar, mirando perpendicularmente a éste. Los focos se situaron a la misma altura, orientados hacia el punto de enfoque de la cámara. Las capturas fueron almacenadas en formato JPG, con una resolución de 4608×2592 píxeles, una densidad de 180 píxeles por pulgada y una profundidad de color de 24bits. En la Figura 1 se muestra, a modo de ejemplo, una de las capturas tomadas en laboratorio.

Figura 1: Ejemplo de imagen capturada para uno de los conjuntos de aceitunas bajo estudio.

Una vez finalizada la captura de las imágenes de los conjuntos, se realizó el pesado individualizado (en gramos, g) de cada fruto usando la balanza de precisión KERN PCB 3500-2 (KERN & Sohn GmbH, Balingen, Alemania). También se realizaron las mediciones de los diámetros mayores y menores (en milímetros, mm) de cada individuo, utilizando un calibrador digital, con 0.01mm de resolución y 0.02mm de precisión. Finalmente, cada medición se asoció unívocamente, mediante etiquetado, a su correspondiente aceituna fotografiada.

2.2 ALGORITMO DE ANÁLISIS DE IMAGEN

La metodología propuesta se basa en la extracción automática de características de las imágenes que son descriptivas del peso y calibre de las aceitunas. Para ello, el algoritmo desarrollado emplea técnicas basadas en morfología matemática y segmentación por umbralización.

2.2.1 Preprocesamiento

En primer lugar, las imágenes son escaladas a un 40% de su tamaño original mediante interpolación bicúbica, ello para reducir la carga computacional de los algoritmos. Seguidamente, se aplica un filtrado gaussiano para la atenuación de ruido de sal y pimienta, con kernel de 5×5 y desviación típica de 0.8. Finalmente, las imágenes se transforman, desde su espacio de color nativo RGB, a una representación en base al espacio de color HSV (del inglés *Hue*, *Saturation*, *Value*). Esta elección estuvo motivada por las características propias de las imágenes capturadas en laboratorio, así como por la propia naturaleza del experimento. Y es que, en efecto, abordar una segmentación basada en características de color carecía de sentido aquí, al no existir garantías de homogeneidad entre las distintas aceitunas en cuanto

a matiz (hecho especialmente destacable en la variedad Arbequina). Sin embargo, el utilizar un tapete de color blanco como fondo de escena propició que sus píxeles presentaran *valores de color* o *luminosidad* considerablemente mayores a los de los píxeles pertenecientes a las aceitunas. Por todo ello, se decidió la conversión de las imágenes al espacio HSV, que proporciona la información de luminosidad, en el canal *V*, desacoplada de otras características. Cabe mencionar que hay más espacios de color potencialmente válidos para este trabajo por ofrecer la misma característica, como por ejemplo los CIELAB (canal *L*) [2] o HSL (canal *L*) [11], entre otros.

2.2.2 Segmentación de las aceitunas

Una vez ejecutada la transformación al espacio HSV, el canal correspondiente a la información del valor de color (*value* en inglés) es almacenada en una matriz independiente, y tratada a partir de este punto como una imagen en escala de grises que, tras diversas transformaciones, será finalmente la imagen binarizada.

Para la matriz contenedora de dicha componente, se realiza una inversión de los valores de sus elementos en base al máximo valor *de gris* contemplado, esto es 255. Así, sea *V* la imagen definida en el intervalo [0,…, 255] (8 bits) correspondiente al canal *Value* del espacio de color HSV, la imagen V_{INV} es la que resulta de la siguiente operación:

$$V_{INV} = 255 - V \tag{1}$$

Con esto, el objetivo perseguido es que los valores de gris mayores (más próximos al blanco) sean los propios de los píxeles que se corresponden con las aceitunas, y que, por tanto, el fondo de la imagen pase a ser la parte más oscura de ésta.
A continuación, para obtener una estimación de fondo, se realiza una apertura morfológica sobre la matriz resultante de la inversión de valores anterior, V_{INV}:

$$I_\gamma = \gamma_\beta(I_{INV}) = \delta_\beta(\varepsilon_\beta(I_{INV})) \tag{2}$$

donde β es un elemento estructurante circular de radio 50 píxeles (lo suficientemente grande como para contener completamente una aceituna de la imagen), y δ y ε las operaciones morfológicas básicas de dilatación y erosión, respectivamente [10].
En búsqueda de una imagen de alto contraste entre las aceitunas y el fondo, la estimación de fondo calculada es sustraída a la matriz de la que origina:

$$I_C = I_{inv} - I_\gamma \tag{3}$$

donde I_C es la imagen de *alto contraste*.
Como última transformación previa a la segmentación definitiva, se aplica una operación de relleno por difusión para incrementar el valor de aquellas regiones más oscuras que aparecen sobre las aceitunas, y que se corresponden con los puntos de máximo brillo, consecuencia del fenómeno de reflectividad máxima de la luz sobre la superficie convexa de los frutos. Complementariamente, y debido al procesamiento previo aplicado, este operador promueve también la homogeneización de los valores de gris del fondo de la imagen. Matemáticamente:

$$I_F = R^\varepsilon_{I_C}(I_m), \tag{4}$$

donde R es la operación de reconstrucción morfológica, consistente en la erosión (ε) iterativa de la imagen I_m con respecto a I_C, con un elemento estructurante unitario, hasta la idempotencia [10]:

$$R^\varepsilon_{I_C}(I_m) = \varepsilon^i_{I_C}(I_m), \text{ siendo} \tag{5}$$

$$I_m(x,y) = \begin{cases} I_C(x,y) & si\ (x,y) \in borde\ de\ I_C \\ \max(I_C) & en\ otros\ casos \end{cases}, \text{ y}$$

$$\varepsilon^i_{I_C}(I_m) = \varepsilon^{i+1}_{I_C}(I_m), \qquad \varepsilon^1_{I_C}(I_m) = \varepsilon_{\beta=1}(I_m) \vee I_C$$

Finalmente, la imagen resultante I_F, aún en escala de grises, es sometida al proceso de binarización por umbralización. Para ello, se establece como umbral su mínimo valor de gris, el cual corresponde a los píxeles pertenecientes al fondo (como ya se ha comentado, el propósito de algunas de las transformaciones realizadas es el de conseguir la uniformidad en cuanto a valor para dichos píxeles). Así, los valores por encima de dicho mínimo se asocian con píxeles de aceituna y se segmentan del fondo (ver Figura 2):

$$I_{BIN}(x,y) = \begin{cases} 255 & si\ I_F(x,y) > \min(I_F) \\ 0 & en\ otros\ casos \end{cases} \tag{6}$$

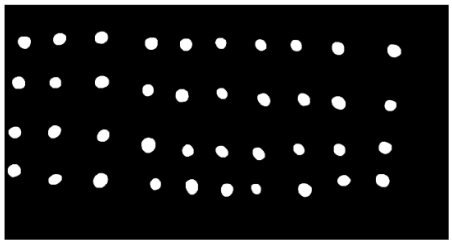


Figura 2: Resultado de la segmentación de la imagen mostrada en la Figura 1.

2.2.3 Postprocesamiento

Como último paso, se aplica una apertura morfológica con un elemento estructurante pequeño para eliminar aquellas componentes conexas con un área (número de píxeles) anormalmente pequeña como para corresponder a una aceituna segmentada:

$$I_{DEF} = \gamma_\beta(I_{BIN}), \tag{7}$$

donde β es un elemento estructurante circular de 10 píxeles de radio.

A continuación (Figura 3), se ilustra el proceso de segmentación a través de los resultados obtenidos en sus distintas etapas, para una sub-imagen de una de las capturas bajo estudio:

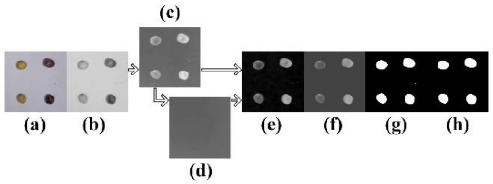


Figura 3: proceso de segmentación para una sub-imagen de una de las capturas bajo estudio: (a) imagen original; (b) canal de valor; (c) canal de valor invertido; (d) estimación de fondo; (e) substracción del fondo estimado; (f) homogeneización de valores; (g) segmentación; (h) imagen final tras postprocesado.

2.2.4 Evaluación de la calidad de la segmentación

Se realizó una segmentación a mano para cada una de las capturas, utilizando un editor de gráficos, con el objetivo de obtener un resultado lo más cercano posible a una segmentación ideal (*Gold Standard*). El conjunto de imágenes de referencia resultante fue utilizado para, mediante comparación con las generadas por el algoritmo, determinar la calidad de la segmentación proporcionada por éste, a través del cómputo de las métricas *RC* (*Recall*) y *PC* (*Precision*), propuestas a continuación:

$$RC = \frac{TP}{TP+FN} \qquad PC = \frac{TP}{TP+FP}, \quad (8)$$

dónde *TP* (*True Positives*), se refiere al número de píxeles categorizados por el algoritmo como pertenecientes a una aceituna (son píxeles blancos en la imagen binarizada) que también lo son según la segmentación ideal (*Gold Standard*). *FN* (*False Negatives*), se corresponde con el número de píxeles erróneamente clasificados como no pertenecientes a alguna aceituna. De la misma forma, FP (*False Positives*), se refiere al número de píxeles considerados erróneamente como pertenecientes a alguno de los frutos, en base a la clasificación propuesta por la segmentación *Gold Standard*.
Junto a *RC* y *PR*, se propone una tercera métrica, $F_{measure}$, basada en los mismos principios, y que evalúa la calidad global de la segmentación:

$$F = 2\,\frac{PC\ x\ RC}{PC+RC} \quad (9)$$

2.3 EXTRACCIÓN DE INFORMACIÓN Y DESARROLLO DEL MODELO DE ESTIMACIÓN

Una vez generadas las distintas segmentaciones, es necesario extraer de éstas la información necesaria para realizar un estudio de correlación con las mediciones de peso y calibre obtenidas previamente en laboratorio. Dentro de cada imagen binaria, para cada región o componente conexa correspondiente a un fruto, se calcula el número de píxeles que la integran; éste es el valor a enfrentar con el peso, medido en gramos, de la aceituna correspondiente. Por otra parte, se determina la elipse que mejor encaja o inscribe la correspondiente componente; esto es, la elipse con el mismo segundo momento central normalizado que dicha región. Los diámetros de dicha elipse, medidos en píxeles, son los datos a enfrentar con las mediciones, en mm, realizadas para los ejes mayor y menor de las aceitunas.
Una vez es adquirido este conjunto de datos, para cada una de las dos variedades consideradas en el experimento, se procede como sigue: del conjunto de 200 frutos que comprende la muestra de estudio, se toman 50, representando la variabilidad del conjunto de la muestra, para conformar un conjunto de entrenamiento. Mediante un análisis de regresión, basado en la relación de dependencia entre los pesos de las aceitunas de dicho conjunto y el número de píxeles que corresponden a las mismas (según lo aportado por la segmentación), se obtiene la recta de regresión que determina, mediante la ecuación que la describe, el modelo predictivo obtenido. Este modelo es aplicado sobre los 150 frutos restantes, actuando dicho subgrupo como conjunto de validación externo, a partir del cual se evalúa la calidad del modelo. El proceso es replicado para los otros parámetros bajo estudio, esto es, para los ejes mayor y menor de los frutos.

Insistiendo en que este procedimiento es realizado de forma independiente para cada una de las variedades, en última instancia, se contempla la generación de un único modelo válido para ambas de forma conjunta. Para ello, se hace uso de los dos conjuntos de entrenamiento establecidos anteriormente, abordando el análisis de regresión a partir de la unión de ambos, y aplicando el resultado sobre la suma de los conjuntos de validación externa correspondientes. Esto, tanto para el peso cómo para la longitud de los ejes que determinan el calibre de los frutos.

2.3.1 Evaluación de la calidad de los modelos de estimación

Una vez calculados los modelos de predicción, y aplicados éstos sobre las poblaciones correspondientes, se llevó a cabo una evaluación final de las estimaciones realizadas. Para ello se recurrió al

cálculo de la raíz del error cuadrático medio, según lo establecido a continuación:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2}{n}} \quad (10)$$

Donde $\widehat{y_i}$ y y_i son, respectivamente, el valor predicho y el valor real del parámetro a estimar, para el fruto *i-ésimo* de la muestra tratada, de un total de *n* elementos.

También se realizó el cálculo del error cuadrático relativo a la media como porcentaje, en base a la siguiente definición:

$$SE = \frac{RMSE}{\frac{\sum_{i=1}^{n} y_i}{n}} \; x\; 100 \quad (11)$$

Finalmente, se determinó el error relativo (como porcentaje), una vez más para cada modelo y la población de aceitunas correspondiente. Dicho error se calculó según se indica a continuación:

$$|E| = \frac{|\sum_{i=1}^{n}(\widehat{y_i} - y_i)|}{\sum_{i=1}^{n} y_i} \; x\; 100 \quad (12)$$

3 RESULTADOS Y DISCUSIÓN

3.1 RESULTADOS DE LA EVALUACIÓN DE CALIDAD DE LA SEGMENTACIÓN

Como se ha comentado anteriormente, la viabilidad del algoritmo para la segmentación ha sido testada en base a la comparación de las imágenes generadas por éste con segmentaciones *presumiblemente* ideales (*Gold Standard*), generadas a mano. Esto, a través de las métricas propuestas en el apartado anterior. En la Tabla 1 se muestran los resultados obtenidos, mostrando los valores para cada una de las segmentaciones realizadas. Según los datos recogidos en ella, se puede concluir la solidez del algoritmo de segmentación propuesto. Cabe destacar que no se observan importantes diferencias entre las distintas variedades, manteniéndose los resultados obtenidos siempre dentro de unos márgenes similares. Esta uniformidad respalda la decisión de no basar el proceso de segmentación en características de color (dadas las significativas diferencias en cuanto a tono entre las dos variedades bajo estudio), y sugiere su correcto funcionamiento tratando nuevas variedades, o frutos en distintos estados de maduración. Otra de las bondades aportadas por el método es su eficacia a la hora de detectar únicamente aquellas componentes conexas que se corresponden con aceitunas, descartando en la segmentación final cualquier componente anómala que no pueda corresponderse con algún fruto. Esto hace posible determinar de forma exacta el número de frutos que aparecen en cada imagen.

Tabla 1: Resultados de la evaluación de la calidad y precisión de las segmentaciones realizadas por el algoritmo propuesto.

muestra	*RC*	*PC*	$F_{measure}$
A1	1	0,8633652	0,9266731
A2	1	0,87684168	0,93438002
A3	1	0,88152872	0,93703456
A4	1	0,89458439	0,9443595
A5	1	0,8669487	0,92873329
promedio	1	0,87665374	0,93423609
P1	0,99997252	0,87503607	0,93334188
P2	1	0,89381748	0,94393202
P3	0,99999426	0,88897784	0,94122379
P4	1	0,90662967	0,9510286
promedio	0,99999336	0,89111527	0,94238157

3.2 RESULTADOS DE LA EVALUACIÓN DE CALIDAD DE LOS MODELOS DE PREDICCIÓN

La correlación existente entre la información extraída a partir de las segmentaciones (las áreas como número de pixeles de las componentes conexas que pertenecen a los frutos, y la longitud, también en píxeles, de los diámetros de las elipses que las inscriben) y las mediciones de peso y calibre de las aceitunas correspondientes, puede observarse en los gráficos mostrados en las figuras 4, 5 y 6. Nótese los altos valores de correlación obtenidos.

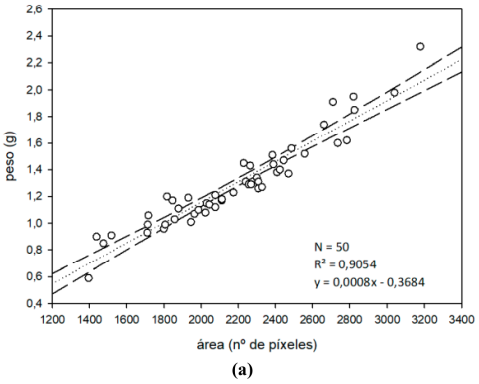
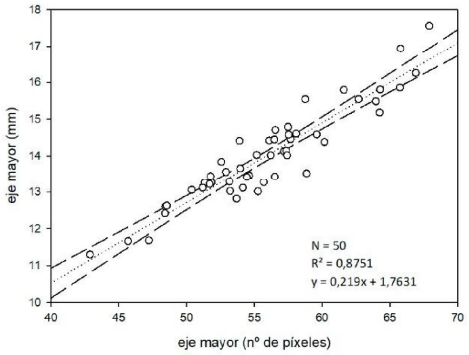
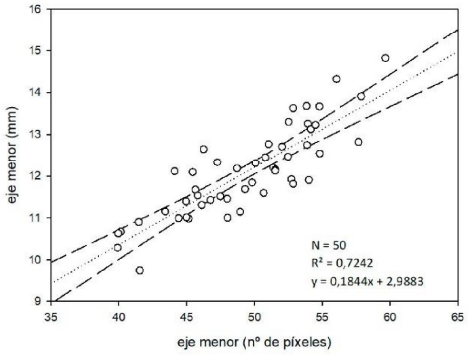


Figura 4: Obtención de los modelos de predicción de la variedad Arbequina a partir del conjunto de entrenamiento, para los parámetros bajo estudio: (a) peso.

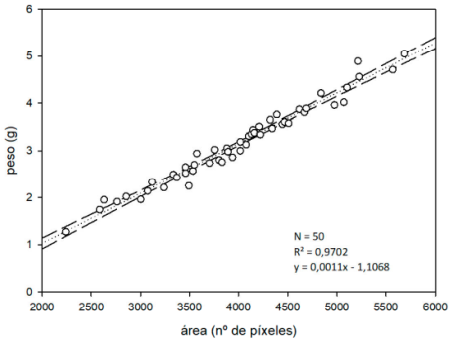


(b)

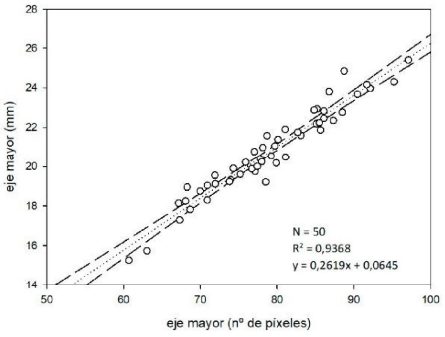


(c)

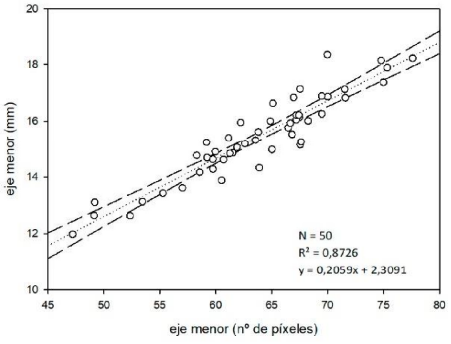
Figura 4: Obtención de los modelos de predicción de la variedad Arbequina a partir del conjunto de entrenamiento, para los parámetros bajo estudio: (b) eje mayor; (c) eje menor.



(a)

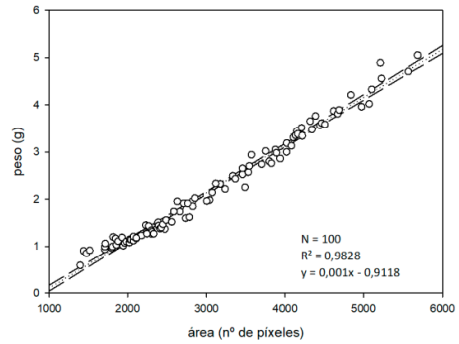


(b)

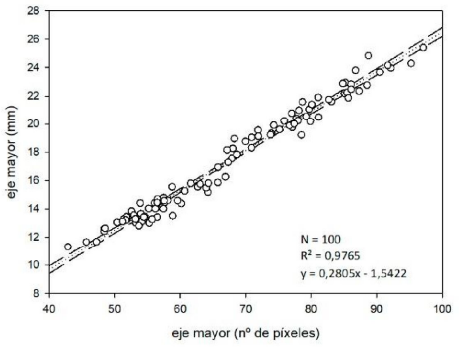


(c)

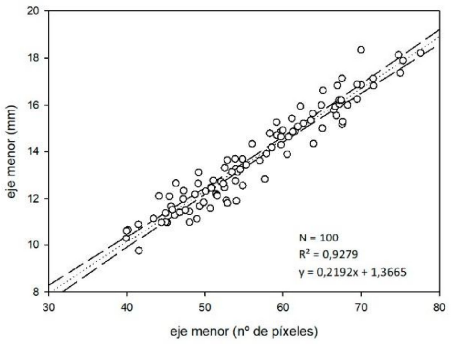
Figura 5: Obtención de los modelos de predicción de la variedad Picual a partir del conjunto de entrenamiento, para los parámetros bajo estudio: (a) peso; (b) eje mayor; (c) eje menor.



(a)



(b)



(c)

Figura 6: Obtención de los modelos de predicción independientes de la variedad a partir de la unión de los dos conjuntos de entrenamiento, para los parámetros bajo estudio: (a) peso; (b) eje mayor; (c) eje menor.

Por otra parte, con el objeto de ilustrar la calidad predictiva de los modelos, se muestran en la Tabla 2 los resultados de su aplicación sobre los conjuntos de validación externa, para cada una de las características bajo estudio.

Tabla 2: Resultados de las distintas mediciones computadas para evaluar la calidad de los modelos de estimación, tras ser estos aplicados sobre los conjuntos de validación externa correspondientes (según variedad), para cada una de las características bajo estudio: peso (a), eje mayor (b) y eje menor.

PESO			
Variedad	*RMSE (g)*	*SE (%)*	*|E| (%)*
Arbequina	0,131495	10,36649	7,406528
Picual	0,256875	8,105015	5,602251
Ambas variedades (modelo global)	0,154839	6,97819	1,105803

(a)

EJE MAYOR			
Variedad	*RMSE (mm)*	*SE (%)*	*|E| (%)*
Arbequina	0,538303	3,81445	0,756152
Picual	0,505237	2,407021	0,325996
Ambas variedades (modelo global)	0,580499	3,307466	0,060863

(b)

EJE MENOR			
Variedad	*RMSE (mm)*	*SE (%)*	*|E| (%)*
Arbequina	0,557464	4,635956	0,310614
Picual	0,608311	3,917913	0,04032
Ambas variedades (modelo global)	0,62158	4,51218	0,117851

(c)

Estos resultados ponen de relieve la viabilidad del método, esto es, la estimación del peso y calibre de distintas aceitunas capturadas en una imagen a partir de la información extraída a partir de una segmentación de ésta, y por extensión, el uso del análisis de imagen para dar solución al problema planteado.

Atendiendo a que la calidad de la segmentación para ambas variedades es similar, las leves diferencias en los valores de error cometido en las estimaciones de peso, así como los mejores resultados de correlación para el caso de las aceitunas picuales, se pueden explicar por las propias diferencias morfológicas existentes entre los frutos de estas variedades. El método aproxima el problema mediante la proyección de la aceituna en un plano, y considera que cada píxel tiene una aportación equitativa en su peso. Esta aproximación, aunque válida a tenor de los resultados obtenidos, puede empobrecer éstos ligeramente para según que variedades, en base a determinadas características morfológicas, tales como su esfericidad. En el caso de los parámetros que determinan el calibre de los frutos, el no considerar estas características también puede tener cierto impacto en la precisión con que se realizan las mediciones de los mismos. Como se ha comentado anteriormente, estas mediciones de los ejes mayor y menor de cada fruto se realizaron utilizando un

calibrador digital. Esta forma de abordar dicha tarea introduce inevitablemente el error humano como factor a considerar en el proceso, que es mayor aún dada la naturaleza y características de los frutos a medir, dada su falta de firmeza. Pero el hecho interesante a destacar es el de que, si bien es cierto que identificar el eje o diámetro de simetría mayor de cada uno de los frutos, y realizar su medición, es un trabajo relativamente sencillo (al menos con las dos variedades tratadas), determinar el eje menor no es una tarea obvia, y su medición está sujeta a un mayor grado de incertidumbre. Este hecho podría explicar las diferencias en los resultados, así como el mejor comportamiento de los modelos computados para el eje mayor respecto a los que estiman el diámetro menor.

Por otro lado, es interesante destacar la ausencia de indicios de un beneficio claro en el uso de modelos específicos para cada variedad en detrimento de un modelo general. Ello potencia la usabilidad y generalización de la metodología desarrollada.

Finalmente, cabe también reseñar que no hay evidencias claras que indiquen la necesidad de explorar modelos no lineales más complejos y de más difícil manejo.

4 CONCLUSIONES

En el presente estudio, se ha desarrollado un método para la estimación del peso y calibre de aceitunas mediante análisis de imagen. Los resultados obtenidos muestran la solidez y precisión de éste, y respaldan el uso de este tipo de técnicas en la implementación de sistemas de calibrado y selección, basados en visión artificial, dentro la industria olivícola.

Cabe destacar la estabilidad en el comportamiento del método al tratar con dos tipos distintos de aceitunas y con distintos estados de maduración, hecho que pone de relieve sus posibilidades como estimador independiente de la variedad. Por otra parte, en todas las imágenes testeadas, el número de componentes conexas segmentadas por el algoritmo coincidió exactamente con el número de frutos. Esta eficacia apoya la confianza en el uso de la solución presentada también para realizar un control preciso del número de aceitunas procesadas en un hipotético sistema real basado en ella.

El presente trabajo admite cierto margen para ser extendido en investigaciones futuras. Así, resultaría interesante cuantificar el impacto de la ponderación de la contribución de los píxeles de aceituna en la estimación de su peso. También es necesario corroborar los resultados obtenidos sobre un conjunto de datos más extenso, y considerando un mayor número de variedades de aceituna, al objeto verificar su generalidad.

Agradecimientos

Este trabajo se ha realizado en el contexto del proyecto TecnOlivo, aprobado en la primera convocatoria del Programa Interreg V-A España - Portugal (POCTEP) 2014-2020, y cofinanciado con fondos FEDER.

English summary

AUTOMATED OLIVE-FRUIT WEIGHT AND SIZE ESTIMATION BY MEANS OF IMAGE ANALYSIS

Abstract

The sizing and sorting of agricultural commodities is a high relevance activity in food industry. This study, focused on the olive farming sector, presents a solution based on image analysis which allows the automatic and non-invasive estimation of the weight and size (major and minor axis) of a set of olive fruits. Considering two different varieties of olive fruits (Arbequina and Picual), a segmentation algorithm, able to extract from images the needed information to compute the weight and size prediction models, was developed. The effectiveness of the proposed method was assessed by calculating the root-mean-square error (RMSE) produced by the models when applied to the corresponding external validation sets. The measured results show evidences of viability as a base to the development of a low-cost olive fruit grading system based on machine vision.

Keywords: olive fruit, weight, size, segmentation, machine vision, estimation.

Referencias

[1] Benalia, S., Bernardi, B., Cubero, S., Leuzzi, A., Larizza, M., Blasco, J., 2015. Preliminary trials on hyperspectral imaging implementation to detect mycotoxins in dried figs. Chem. Eng. Trans. 44, 157–162.

[2] Connolly, C., Fliess, T. A, 1997. Study of efficiency and accuracy in the transformation from RGB to CIELAB color space. IEEE Trans. Image Process. 6, 1046–1047.

[3] Gómez-Sanchis, J., Blasco, J., Soria-Olivas, E., Lorente, D., Escandell-Montero, P., Martínez-Martínez, J.M., Aleixos, N., 2013. Hyperspectral LCTF-based system for classification of decay in mandarins caused by Penicillium digitatum and Penicillium italicum using the most relevant bands and non-linear classifiers. Postharvest Biol. Technol. 82, 76–86.

[4] International Olive Oil Council (IOC), 2018. IOC updates series of world statistics on production, imports, exports and consumption. Available: http://www.internationaloliveoil.org/estaticos/view/131-world-olive-oil-figures.

[5] Kailis, S.; Harris, D. Producing Table Olives; Publisher: Landlinks Press, Collingwood, Vic., 328, Australia, 2007; pp. 173-174, pp. 244-246

[6] Moreda, G. P.; Ortiz-cañavate, J.; García-ramos, F. J.; Ruiz-altisent, M. Non-destructive technologies for fruit and vegetable size determination – a review. J. Food Eng. 2009, 92, 119–136

[7] Pallottino, F., Costa, C., Antonucci, F., Menesatti, P., 2013a. Sweet cherry freshness evaluation through colourimetric and morphometric stem analysis: two refrigeration systems compared. Acta Aliment. 42, 428–436.

[8] Pallottino, F., Menesatti, P., Lanza, M.C., Strano, M.C., Antonucci, F., Moresi, M., 2013b. Assessment of quality-assured Tarocco orange fruit sorting rules by combined physicochemical and sensory testing. J. Sci. Food Agric. 93 (1176), 1183.

[9] Sánchez Gómez, A. H.; García García, P.; Rejano Navarro, L. Elaboration of table olives. Grasas y Aceites 2006, 57, 86–94.

[10] Soille, P. (2013). Morphological image analysis: principles and applications. Springer Science & Business Media, Germany.

[11] Sonka, M., Hlavac, V., Boyle, R, 2014. Image processing, analysis, and machine vision. Cengage Learning, USA.

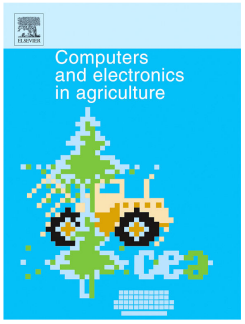
[12] United States Department of Agriculture, Foreign Agricultural Service. Oilseeds: World Markets and Trade. Available online: https://apps.fas.usda.gov/psdonline/circulars/oilseeds.pdf (accessed on 19 June 2018).

4.2. Preharvest in-field yield estimation

4.2.1. Article 4

Identification of olive fruit, in intensive olive orchards, by means of its morphological structure using convolutional neural networks

A. Aquino, J. M. Ponce, J. M. Andújar



Published in:

Journal: Computer and Electronics in Agriculture (ISSN: 0168-1699)

Editorial: Elsevier

Reference: vol. 176, p. 105616

Year: 2020

DOI: 10.1016/j.compag.2020.105616

Quality index (Journal Citation Reports®, 2019): 24/109 (Q1) in the category "Computer Science, Interdisciplinary Applications". Impact Factor of 3.858.

4.2.2. Congress 4

IDENTIFICACIÓN Y CONTEO DE ACEITUNAS EN IMÁGENES DIGITALES TOMADAS EN EL OLIVAR MEDIANTE MORFOLOGÍA MATEMÁTICA Y REDES NEURONALES CONVOLUCIONALES

A. Aquino, J. M. Ponce, B. Millan, D. Tejada, J. M. Andújar

Published in:

XL JORNADAS DE AUTOMÁTICA. Ferrol, Galicia, Spain, September 4-6, 2019. PROCEEDINGS. (ISBN: 978-84-9749-716-9)

Editor: José Luis Calvo Rolle [et al.]

Reference: pp. 818-827

Year: 2019

XL Jornadas de Automática Visión por Computador

IDENTIFICACIÓN Y CONTEO DE ACEITUNAS EN IMÁGENES DIGITALES TOMADAS EN EL OLIVAR MEDIANTE MORFOLOGÍA MATEMÁTICA Y REDES NEURONALES CONVOLUCIONALES

Arturo Aquino, Juan Manuel Ponce, Borja Millan, Diego Tejada-Guzmán, José Manuel Andújar
{arturo.aquino, jmponce.real, borja.millan, diego.tejada, andujar}@diesia.uhu.es
Departamento de Ingeniería Electrónica, Sistemas Informáticos y Automática. Universidad de Huelva.
Carretera Huelva-Palos, s/n, 21810 Palos de la Frontera (España).

Resumen

La estimación precoz y precisa de la producción es un objetivo muy codiciado en la agricultura moderna. En el caso de la olivicultura, ello toma una especial relevancia debido al alto valor económico que alcanza su producción. Este artículo presenta una metodología enfocada a lograr dicho objetivo. Concretamente, se propone un algoritmo de visión artificial capaz de detectar las aceitunas visibles en una imagen digital de un árbol de olivo, tomada directamente en campo, de noche y con iluminación artificial. En primera instancia, esta imagen es preprocesada mediante técnicas de morfología matemática y filtrado estadístico para, a partir de ella, obtener un conjunto de subimágenes con alta probabilidad de contener una aceituna. Este preprocesamiento reduce el espacio potencial de búsqueda en una magnitud de 10^3. A continuación, estas subimágenes son clasificadas por una red neuronal convolucional como 'aceituna' o 'descarte'. De un total de 304.483 subimágenes, extraídas de 21 imágenes, la red clasificó correctamente el 98,23%, y arrojó un coeficiente de determinación R^2 igual a 0,9875, al enfrentar el número de aceitunas detectadas con el obtenido manualmente. Esta precisión alcanzada indica que el algoritmo desarrollado constituye un paso certero en la implementación de un futuro sistema de estimación de la producción de cultivos de olivo.

Palabras clave: Agricultura de precisión, estimación de la producción, aceituna, visión artificial, red neuronal convolucional.

1 INTRODUCCIÓN

El cultivo del olivo (*Olea europaea* L.), y su mercado asociado, es un notable motor económico para buena parte de la cuenca mediterránea. Lo es especialmente en el caso de España, la cual aportó en 2017 el 31,38% del total de aceitunas producidas mundialmente [1]; esta cifra asciende hasta el 61,95% cuando se considera también la producción de Portugal, Italia y Grecia [1]. Además, el cultivo del olivo trasciende de lo económico a lo social, ya que, por ejemplo, el aceite de oliva es la piedra angular de la dieta mediterránea, la cual ha sido declarada Patrimonio Inmaterial de la Humanidad por la UNESCO [2].

La agricultura en general, y la olivicultura en particular, se enfrenta a importantes desafíos, como son la sostenibilidad medioambiental y la productividad [3]. Estas dificultades, lejos de estar cerca de ser solventadas, se recrudecerán al ritmo exponencial de crecimiento esperado de demanda de alimentos, ligado éste a las estimaciones de desarrollo de la población mundial (según estima la ONU, desde los actuales 7.600 M de habitantes, se llegará a los 9.800 M en 2050 [4]).

En este contexto, es donde la agricultura de precisión recibe un apoyo y una atención muy relevantes en los últimos años [5]. Tradicionalmente, los cultivos han sido gestionados gracias a la capacidad de percepción de los agricultores para evaluar su desarrollo vegetativo, su necesidad de riego o nutrientes, la madurez del fruto, la cantidad de cosecha esperada, etc. En contraposición, la agricultura de precisión auspicia el uso de tecnología para adquirir información objetiva y de alta precisión espacio-temporal acerca del estado de los cultivos, para posibilitar así una óptima gestión de los mismos [5].

De entre la información deducible del olivar, la estimación temprana y precisa de la cosecha destaca en interés, ya que sería una valiosa herramienta de apoyo al sector [6]. Y es que, en efecto, una estimación precisa de la producción tendría aplicaciones prácticas en muy diversos aspectos, tales como: eficiencia en la transformación del aceite de oliva, gestión del stock, o la optimización de los recursos humanos necesarios para la recolección [7]. También tendría una importancia crucial en la regulación del precio de mercado del aceite de oliva. La volatilidad de este mercado está fuertemente ligada a las expectativas de producción, actualmente evaluadas de manera visual.

https://doi.org/10.17979/spudc.9788497497169.818

Ello ha llevado frecuentemente a unos niveles de inestabilidad de precios declaradamente excesivos [8]. La importancia de la estimación temprana de la cosecha es transversal a todos los cultivos con productos derivados de alto valor de mercado. Como ejemplo valga el cultivo de vid, para el que se pueden encontrar en la bibliografía una variedad de metodologías experimentales basadas en visión artificial enfocadas a ello [9-12]. Otros cultivos como el de la palma [13], el arroz [14], los cereales [15,16], o los frutales [17-19], han recibido también una atención relevante a este respecto. Sin embargo, para el cultivo del olivo, la bibliografía es parca en estudios y, en su mayoría, están basados en estimar la producción a través de variables indirectas. Ejemplo de ello son los trabajos referenciados en [20-22], los cuales plantean la resolución del problema a través de la generación de modelos alimentados con variables meteorológicas y de polinización. La principal debilidad de este enfoque radica en la naturaleza misma de estas variables, con un fuerte componente estocástico, lo cual compromete la reproducibilidad de los modelos. Además, el fenómeno de formación y mantenimiento del fruto es virtualmente muy complejo y resultado de la conjunción de multitud de variables. A este respecto, los trabajos referenciados no demuestran la representatividad de las variables seleccionadas con relación a las omitidas.

Este artículo presenta una metodología basada en visión artificial para la detección y conteo de las aceitunas visibles en imágenes digitales tomadas en campo. Para ello, primero se genera, a partir de las imágenes, un conjunto reducido de candidatos (subimágenes) mediante operadores de morfología matemática, reduciendo así el espacio potencial de búsqueda en una magnitud de 10^3. A continuación, estas subimágenes son preprocesadas y clasificadas por una red neuronal convolucional (CNN), la cual valida o descarta cada instancia en función de si contiene una aceituna o no. El nivel de precisión de la metodología propuesta argumenta a favor de su integración en un futuro sistema para la estimación precoz de la cosecha del olivar.

2 MATERIALES Y MÉTODOS

2.1 ADQUISICIÓN DE IMÁGENES

En septiembre de 2018 (dos meses antes de recolección), se adquirió un conjunto de 36 imágenes de la variedad Picual en un olivar de cultivo intensivo situado en Gibraleón, (37°20'09.2"N 7°02'19.8"W), provincia de Huelva (Andalucía, España); el marco de plantación de la parcela era de 5,5 × 7 m. Los individuos de estudio se seleccionaron para conseguir la máxima variabilidad disponible en cuanto a su capacidad productiva de aceitunas.

Las imágenes se tomaron de noche con la iluminación artificial proporcionada por un foco halógeno de 500 W, ya que ello favorece la simplificación de la escena en el fondo del objeto de interés. Esta decisión fue tomada gracias a la experiencia previa de los autores en trabajos similares realizados en vid [12]. La captura de las imágenes se realizó de manera manual y con trípode, usando para ello una cámara sin espejo Sony α7-II (Sony Corp., Tokyo, Japan) equipada con una óptica Zeiss estabilizada de 24/70 mm (Zeiss Gruppe, Oberkochen, Germany). Dicha cámara monta un sensor CMOS de 24 Mpx estabilizado en cinco ejes, y fue configurada con las siguientes características: apertura de f/14, tiempo de exposición de 1/200 s, distancia focal de 24 mm, resolución de 6.000 x 3.376 píxeles, profundidad de color de 24 bits, y formato de almacenamiento de imagen JPEG de mínima compresión. La distancia a la que se tomaron las imágenes osciló entre los 2 m y los 4 m dependiendo del porte del árbol para, en cada caso, obtener un encuadre apropiado de toda su superficie foliar. La figura 1 muestra un ejemplo de imagen capturada bajo los preceptos descritos.

Figura 1: imagen de olivo de la variedad Picual, sujeto de este estudio, obtenida de noche con iluminación artificial.

2.2 METODOLOGÍA DE ANÁLISIS DE IMAGEN

La metodología se basa en el uso de una red neuronal convolucional (CNN) para detectar las aceitunas presentes en una imagen de olivo como la mostrada en la figura 1. Para ello, previamente se aplica un preprocesamiento dirigido a: (1) mejorar las condiciones de partida de la imagen; (2) reducir el espacio de búsqueda mediante la obtención de un conjunto reducido de subimágenes (candidatos) con alta probabilidad de contener individualmente las aceitunas de la imagen; (3) configurar la composición y contenido de las subimágenes para optimizar el rendimiento de la CNN.

La metodología descrita en este artículo fue implementada usando la plataforma Matlab R2018b (The MathWorks, Inc., Natick, Massachusetts, USA).

https://doi.org/10.17979/spudc.9788497497169.818

2.2.1 Preprocesamiento

Sea la imagen digital *I*, análoga a la mostrada en la figura 1, generada de manera nativa de acuerdo con el espacio de color RGB, y codificada con 8 bits por canal. Esta imagen inicial es transformada al espacio de color CIE 1976 *L*a*b**, ya que ello posibilita el análisis independiente de las componentes de iluminación (canal *L**) y color (canales *a** y *b**) [23]. Así, quedan definidas las imágenes *L*, *A* y *B*, correspondientes a los canales *L**, *a** y *b**, respectivamente, que componen la imagen transformada. Para este trabajo, se descartó el uso de la imagen *A* por ofrecer un contraste de niveles de gris ínfimo en las regiones de interés.

En primer lugar, para reducir ruido de "sal y pimienta", se aplica un filtrado promedio circular, dentro de un kernel de convolución *k* de tamaño 11 × 11, a las componentes *L* y *B*:

$$L_s = L * k;\ B_s = B * k \tag{1}$$

donde el círculo de activación del filtro tiene un diámetro de 11 píxeles. El reducido tamaño de este filtro comparado con el de la imagen, 11 x 11 vs 6.000 x 3.376, profiere cierta tolerancia al establecimiento de su valor para conseguir el objetivo perseguido. A su vez, el promediado circular favorece el "limpiado" de los patrones de iluminación análogos las aceitunas.

A continuación, a partir de L_s, se genera un conjunto de semillas que posteriormente servirá para generar un conjunto de subimágenes o candidatos que permita reducir el espacio de búsqueda. Para ello, cada semilla será una agrupación píxeles vecinos (conocida formalmente como componente conexa, *CC*), que representa la ocurrencia de una región máxima local de iluminación. Gracias a la ley del coseno de Lambert [10], se conoce que una superficie convexa, como la de una aceituna, produce un intenso patrón circular de reflexión de la luz. Por ello, esta operación favorece la localización de las aceitunas en la imagen. Estas regiones máximas se obtienen mediante la aplicación de la transformada morfológica *h-maxima*, la cual encuentra aquellas regiones en la imagen que cumplen tener una elevación menor o igual que el escalar *h*. La transformada, calcula en primer lugar la reconstrucción morfológica, *R*, de la imagen L_s a partir del marcador $L_s - h$ (consultar los detalles del operador reconstrucción en [24]),

$$L_{filt} = R_{L_s}(L_s - h); h = 3 \tag{2}$$

para finalmente extraer las regiones máximas en la imagen L_{filt}, mediante el cálculo de

$$L_{RM} = L_{filt} - R_{L_{filt}}(L_{filt} - 1) \tag{3}$$

y la segmentación de la imagen resultante:

$$L_{RMbin} = \begin{cases} 255 & if\ \ L_{RM}(x,y) > 0 \\ 0 & en\ otro\ caso \end{cases} \tag{4}$$

En situaciones de aplicación análogas, se ha discutido acerca del valor óptimo para el parámetro *h* [12, 25]. Estos trabajos concluyeron que valores en torno al 3 o el 4 eran óptimos, aunque no excluyentes, ya que ligeras variaciones también proporcionaban resultados satisfactorios; para la experimentación descrita en este trabajo se estableció $h = 3$. Finalmente, el conjunto de regiones máximas de iluminación (o semillas) se compone con las componentes conexas, CC_i, existentes en la imagen binaria L_{RMbin}:

$$S_{RM} = \{CC_i \subseteq L_{RMbin}\} \tag{5}$$

La figura 2 ilustra el resultado del cálculo del conjunto de componentes conexas provenientes de regiones máximas de iluminación.

La siguiente operación consiste en obtener un conjunto de subimágenes candidatas centradas en las semillas encontradas, tanto para L_s como para B_s; estos conjuntos se denotarán como S_{L_s} y S_{B_s}, respectivamente. Para ello, para cada componente conexa $CC_i \in S_{RM}$ se extrae, en torno a su centroide, una subimagen de L_s y otra de B_s de tamaño 41 × 41. Así, se cumple que

$$S_{L_s} = \{s_{L_s}^i\}, S_{B_s} = \{s_{B_s}^j\}; 1 \le i,j \le \#(S_{RM}) \tag{6}$$

y que

$$\forall i\ 1 \le i \le \#(S_{RM}),\ s_{L_s}^i\ "deriva\ de"\ CC_i\ \wedge s_{B_s}^i\ "deriva\ de"\ CC_i \tag{7}$$

Debido a la variación en la distancia de fotografiado y al volumen de los olivos, las aceitunas impresionan con muy diversos tamaños en las distintas imágenes. Las dimensiones para las subimágenes fueron establecidas empíricamente para que contuvieran completamente a una aceituna independientemente de su tamaño. Por otro lado, en media, se generaron 14.499 subimágenes a partir de una imagen. Teniendo en cuenta que, para imágenes de resolución 6.000 x 3.376 píxeles, el espacio potencial de búsqueda está compuesto por 20.256.000 candidatos (número de píxeles que la componen), éste quedó reducido en un orden de magnitud de 10^3.

El último paso del preprocesamiento consiste en, una vez obtenidas las subimágenes de los conjuntos S_{L_s} y S_{B_s}, encontrar una combinación de ellas tal que a la postre favorezca el rendimiento de una CNN en la tarea de determinar si una subimagen contiene una aceituna o no. Es más, la red deberá ser capaz de discernir si la subimagen en evaluación está centrada

https://doi.org/10.17979/spudc.9788497497169.818

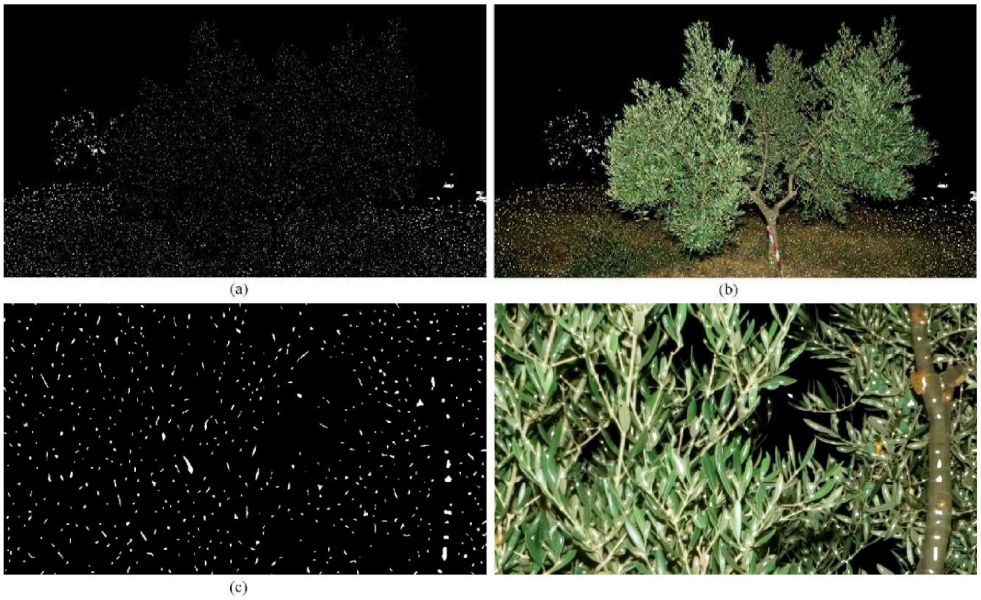


Figura 2: Resultado del procedimiento de obtención de componentes conexas de iluminación, o semillas, para la reducción del espacio de búsqueda: (a) conjunto de semillas obtenidas a partir de la imagen de la figura 1 representadas en blanco sobre fondo negro; (b) conjunto de semillas de (a) representadas sobre la imagen original; (c) sección de la imagen (a); (d) sección de la imagen (b), análoga a la (c).

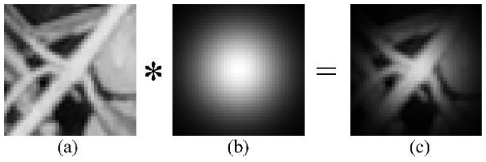


Figura 3: (a) subimagen de L_s, $s_{L_s}^i$, extraída en torno al centroide de una semilla producida por el brillo de una rama; nótese que la imagen contiene representaciones parciales de dos aceitunas; (b) ilustración de una matriz Gaussiana de tamaño 41 × 41 y $\sigma = 9{,}5$; (c) subimagen $s_{L_s'}^i$, resultado de la multiplicación elemento a elemento de (a) y (b), en la que se atenúa la presencia de las aceitunas por no estar centradas en (a).

en un brillo proveniente de una aceituna para validarla, o de cualquier otro elemento, como por ejemplo una hoja, para descartarla. En este sentido, es muy frecuente la existencia de subimágenes que contienen alguna aceituna, total o parcialmente, pero que han sido generadas a partir de un brillo de otro elemento. En este caso, la red deberá de ser capaz de discernir que la subimagen candidata está centrada en un elemento anómalo para descartarla. Para favorecer este precepto, se crea un nuevo conjunto a partir de S_{L_s}, $S_{L'_s}$, en el que sus subimágenes son multiplicadas, elemento a elemento, por una matriz Gaussiana normalizada G de tamaño 41 × 41 y $\sigma = 9{,}5$:

$$S_{L'_s} = \left\{ s_{L'_s}^i \middle| s_{L'_s}^j(k,l) = s_{L_s}^j(k,l) \times G(k,l); \; \forall k \forall l, 1 \le k,l \le 41, 1 \le j \le \#(S_{RM}) \right\} \quad (8)$$

En efecto, la subimagen $s_{L'_s}^i$ contiene la misma información que $s_{L_s}^i$, pero ponderada en importancia de manera decreciente desde el centro hacia el exterior de ésta (ver la figura 3).

Finalmente, la unión de los conjuntos ordenados S_{L_s}, S_{B_s} y $S_{L'_s}$ configura el espacio de búsqueda de la imagen I de la que proceden. Así, cada tripleta de elementos i-ésimos puede ser considerada una subimagen candidata de tres canales y tamaño 41 × 41, en la que $s_{L_s}^i$ aporta información de iluminación, $s_{L'_s}^i$ pondera la relevancia de dicha información, y $s_{B_s}^i$ proporciona conocimiento relativo al color (Figura 4).

2.2.2 Procedimiento de detección de aceitunas

Esta etapa de la metodología de análisis de imagen trata sobre la configuración y el entrenamiento de una CNN capaz de clasificar las subimágenes candidatas generadas en la fase de preprocesamiento, como *aceituna* (caso positivo) o *descarte* (caso negativo).

2.2.2.1 Arquitectura de la CNN

A grandes rasgos, las CNN están compuestas de dos estructuras principales conectadas. La primera,

https://doi.org/10.17979/spudc.9788497497169.818

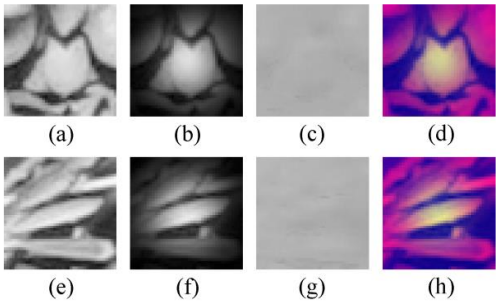


Figura 4: (a)-(d) subimágenes $s_{L_s}^i$, $s_{L_s'}^i$, $s_{L_B}^i$ y representación como imagen RGB de su combinación, respectivamente; son resultado de la semilla *i*-ésima procedente del máximo de iluminación local producido por una aceituna; (e)-(h) subimágenes $s_{L_s}^j$, $s_{L_s'}^j$, $s_{L_B}^j$ y representación como imagen RGB de su combinación, respectivamente; son resultado de la semilla *j*-ésima procedente del máximo de iluminación local producido por una hoja.

encargada de la extracción de características, está en esencia conformada por un conjunto de capas, de dimensionalidad uno o más, que implementan filtros convolucionales, y que están a menudo intercaladas con otras de normalización o reducción del muestreo. Esta primera estructura, a su salida, conecta con un perceptrón multicapa clásico, el cual se encarga de aprender las directrices de clasificación. El valor de todas las variables derivadas de este planteamiento multicapa, es ajustado de forma concurrente e iterativa mediante un algoritmo que persigue la minimización de una función de coste que expresa el error de clasificación de la red. Para una descripción más precisa sobre las CCN, consúltese [26].

Existen arquitecturas que han demostrado gran solvencia en la resolución de problemas complejos, y que han contribuido enormemente al avance del aprendizaje profundo. Entre las más relevantes, consideradas hoy prácticamente un estándar, cabe destacar: AlexNet [27], VGG-16 [28], GoogLeNet/Inception [29], ResNet [30] o Inception-ResNet [31], que hibrida las arquitecturas Inception y ResNet. Debido a la complejidad de las imágenes y las diferentes magnificaciones de los objetos que contienen, se seleccionó Inception-ResNet para el trabajo que se presenta. Por un lado, el enfoque Inception postula la aplicación de convoluciones de distintos tamaños en la misma capa, lo cual aporta flexibilidad en la detección de patrones que pueden aparecer con muy diversas dimensiones. Estos postulados impresionan favorecer el reconocimiento de aceitunas, independientemente de su tamaño. Por su lado, el paradigma ResNet aporta a la red la potencia del tratamiento de los residuos de aprendizaje. Para una capa dada, el residuo es básicamente la diferencia entre lo aprendido al inicio y al final de la misma. La arquitectura ResNet explota este residuo en su misma topología para favorecer y optimizar la convergencia a la solución óptima.

Inception-Resnet, concretamente su versión 2 que fue la usada en este trabajo, es una red de 164 capas y 55,9 M de parámetros. Ello, junto con su arquitectura híbrida anteriormente discutida, le confiere una gran potencia de razonamiento. A su entrada, la red admite imágenes de tamaño 299 × 299 × 3.

2.2.2.2 Entrenamiento y optimización de la CNN

Las CNN enumeradas anteriormente, y algunas otras populares, están disponibles a través de internet, entrenadas con millones de imágenes, para ser manipuladas o usadas con distintas plataformas de desarrollo. Aunque las versiones disponibles no hayan sido entrenadas para el caso de interés requerido, el conocimiento con el que se proporcionan es de gran valor ya que posibilita efectuar lo que se conoce como "transfer learning". Este paradigma consiste en sustituir el perceptrón de la CNN seleccionada por uno configurado y sin entrenar, conservando el bloque convolucional. De esta manera, se parte para el entrenamiento deseado de una CNN con gran capacidad adquirida de extracción de características. Este fue el enfoque empleado en esta investigación.

Previamente a poder realizar el entrenamiento por transferencia de aprendizaje de la CNN, hubo que crear un conjunto de instancias de entrenamiento. Cada instancia estuvo constituida por la tripleta de subimágenes $s_{L_s}^i$, $s_{L_s'}^i$ y $s_{L_B}^i$ (ver ejemplos en la figura 4, imágenes (a)-(c) y (e)-(g)), y fue catalogada como positiva, o *aceituna*, si éstas estaban centradas en una aceituna, o como negativa, o *descarte*, en caso contrario. Para generar dichas instancias catalogadas, en primer lugar, se seleccionaron 15 imágenes de las 36 totales (las 21 restantes se reservaron para validación externa). A continuación, se calcularon sus componentes conexas provenientes de regiones máximas de iluminación aplicando el preprocesamiento formulado en las ecuaciones (1)-(5). Estas componentes conexas se representaron sobre las imágenes originales tal y como se ilustra en la figura 2 y sobre ellas, usando un software de edición de imagen, se etiquetaron aquellas componentes correspondientes con brillos de aceitunas, independientemente del grado de oclusión que presentaran (ver la figura 5). De esta forma, se etiquetaron un total de 9.575 casos positivos, y 208.848 negativos. En torno a las componentes ya etiquetadas, tanto de los casos positivos como de los negativos, se extrajeron las subimágenes $s_{L_s}^i$, $s_{L_s'}^i$ y $s_{L_B}^i$ correspondientes siguiendo el resto del preprocesamiento especificado por las ecuaciones (6)-

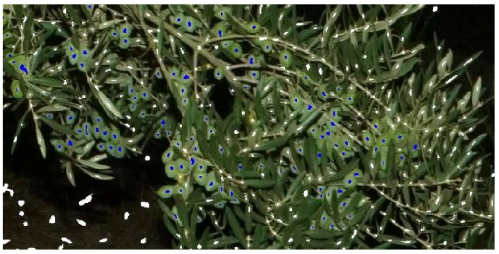


Figura 5: ilustración del etiquetado de instancias de entrenamiento. Se muestra una sección de la imagen de la figura 2-(b) en la que las componentes conexas coincidentes con brillos de aceitunas son etiquetadas en azul como casos positivos. Los casos negativos son etiquetados en color blanco. Nótese que todas las aceitunas, independientemente de su grado de oclusión, son etiquetadas como caso positivo.

(8). Finalmente, para ampliar el conjunto de instancias positivas, las subimágenes implicadas se rotaron 30º, 60º, 90º, …, 330º, para así configurar un conjunto definitivo de instancias positivas de 117.084; de entre las 208.848 instancias negativas, se seleccionó aleatoriamente el mismo número, para obtener así conjuntos balanceados. Finalmente, todas las subimágenes fueron escaladas al tamaño de entrada de la CNN, esto es, 299 × 299, usando interpolación bilineal.

El 80% de las instancias positivas y de las negativas, seleccionadas de forma aleatoria, se usó como conjunto de entrenamiento, mientras que el 20% restante fue empleado en validación para reducir las probabilidades de aparición de sobre-entrenamiento. Como algoritmo de aprendizaje se empleó el del descenso estocástico del gradiente optimizado por momentum (sgdm) [32], con una ratio de aprendizaje que varió desde 10^{-2} hasta 10^{-5} en función de la evolución del proceso de convergencia. Se configuró un tamaño para los mini-conjuntos de entrenamiento de 28 instancias, lo que resultó en 6.690 iteraciones por época. Para llegar a una solución óptima, la CNN entrenó un total de 60 épocas.

3 RESULTADOS Y DISCUSIÓN

3.1 METODOLOGÍA DE EVALUACIÓN DE RESULTADOS

Los resultados de la metodología presentada fueron evaluados utilizando métricas basadas en tablas de contingencia para clasificación binaria. Para ser factible este enfoque, en primer lugar, se generó un conjunto *gold standard* para las 21 imágenes de validación externa. Ello se realizó siguiendo un procedimiento análogo al descrito anteriormente, e ilustrado en la figura 5, para etiquetar las instancias de entrenamiento. Esto es, las imágenes se preprocesaron aplicando las operaciones (1)-(5) para, a continuación, etiquetar manualmente las componentes conexas como *aceituna* o *descarte* usando un software de edición de imagen.

Con la disponibilidad del conjunto *gold standard*, las posibles ocurrencias de clasificación que se podían producir al evaluar las imágenes de validación externa, quedaron definidas como:

- TP (verdadero positivo): instancia catalogada como *aceituna* por la CNN, que también fue etiquetada como tal en el conjunto *gold standard*.
- TN (verdadero negativo): instancia catalogada como *descarte* por la CNN, que también fue etiquetada como tal en el conjunto *gold standard*.
- FP (falso positivo): instancia catalogada como *aceituna* por la CNN, que fue etiquetada como *descarte* en el conjunto *gold standard*.
- FN (falso negativo): instancia catalogada como *descarte* por la CNN, que fue etiquetada como *aceituna* en el conjunto *gold standard*.

Con estas definiciones, las métricas usadas para evaluar el rendimiento del algoritmo de análisis de imagen fueron:

$$PR = \frac{TP}{TP + FP};$$

$$RC = \frac{TP}{TP + FN};$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}; \quad (9)$$

$$F - score = 2 \times \frac{PR \times RC}{PR + RC}$$

Donde *PR* (*precision)* es la tasa de acierto alcanzada en la inferencia realizada sobre la clase *aceituna*, *RC* (*recall*) es la fracción de instancias *aceituna* que el algoritmo logra catalogar, *ACC* (*accuracy*) evalúa la tasa de acierto global alcanzada para las dos clases, *y F-score* es la media armónica de *PR* y *RC*.

Adicionalmente, también se estudió el coeficiente de determinación R^2 resultante de enfrentar, para cada imagen, el número de aceitunas detectadas con el establecido mediante conteo manual (el establecido en el conjunto *gold standard*).

3.2 EVALUACIÓN Y DISCUSIÓN DE RESULTADOS

La tabla 1 detalla, para cada imagen del conjunto de validación externa, el número de aceitunas recogido en el conjunto *gold standard*, el número de instancias generado por el preprocesamiento, y las ocurrencias

Tabla 1: Especificación por imagen perteneciente al conjunto de validación externa, y en total, de las instancias a clasificar generadas por el preprocesamiento, el número de aceitunas reales recogidas en el conjunto *gold standard*, y las ocurrencias de clasificación.

Imagen	Aceitunas reales	Instancias	TP	TN	FP	FN
1	907	13.745	823	11.827	84	188
2	904	14.058	788	12.307	117	58
3	1.225	17.158	981	14.831	244	121
4	702	14.219	544	12.879	158	94
5	1.026	15.973	890	13.907	137	149
6	1.227	14.972	1.113	12.432	114	200
7	882	14.315	794	12.353	88	286
8	795	14.146	727	12.165	69	458
9	984	14.092	891	11.877	93	340
10	965	16.167	862	14.106	103	234
11	368	12.261	299	11.539	69	55
12	560	15.378	469	14.234	91	115
13	740	13.871	646	12.354	94	131
14	566	15.193	507	13.988	59	132
15	493	15.646	432	14.601	61	120
16	543	12.375	483	11.239	60	110
17	939	13.252	846	11.318	94	148
18	938	15.972	842	13.986	97	205
19	969	14.143	866	12.098	105	208
20	75	11.674	25	11.566	50	8
21	63	15.873	38	15.764	25	8
Total	15.871	304.483	13.866	271.371	2.012	3.368

Tabla 2: Resultados de la metodología de análisis de imagen calculados sobre el total de TP, TN, FP y FN producidos para las 21 imágenes de validación externa.

PR	RC	F-score	ACC
0,8733	0,8046	0,8375	0,9823

de clasificación producidas por la CNN; también se incluye esta información agrupando los datos de todas las imágenes como un todo. La tabla 2 muestra los resultados medidos en términos de las métricas definidas anteriormente en (9). Por un lado, y atendiendo al resultado de *Acc*, el rendimiento general del clasificador podría ser calificado de destacable, ya que clasificó correctamente el 98,23% de las instancias. Por otro lado, de acuerdo con el valor de *PR*, se produjo un 87,33% de acierto cuando la CNN clasificó una instancia como "aceituna", mientras que el *RC* indica que el 80,46% de todas las instancias "aceituna" fueron correctamente clasificadas. La conjunción de ambos aspectos, en términos de *F-score*, dio como resultado un rendimiento del 83,75%. Para contextualizar estos resultados, conviene ponderar varias circunstancias. La complejidad de las imágenes para los fines perseguidos fue notable, contribuyendo a esa dificultad: la similitud en color de las aceitunas y las hojas de olivo, la marcada variabilidad en tamaño de las aceitunas, que éstas se presentan tanto aisladas como junto a otras, y que pueden presentar importantes grados de oclusión con otros elementos como hojas, ramas u otras aceitunas. Ninguno de estos distractores fue evadido, ya que se persiguió la detección de aceitunas en cualquier circunstancia, y con cualquier apariencia y grado de oclusión, para lo que se diseñó un preprocesamiento específico que posibilitó alcanzar los resultados descritos.

La figura 6 muestra el resultado del análisis de correlación realizado al enfrentar, por imagen, el número de aceitunas reales recogidas en el conjunto *gold standard*, con las detectadas por la metodología descrita en este artículo. Como se puede apreciar, tanto en la gráfica como en el valor medido de R^2 que alcanzó el 0,9875, el grado de congruencia estadística entre las dos variables fue verdaderamente elevado. En el trabajo presentando en [12] para la estimación de la producción en viñedos, planteado con un enfoque similar, los autores encontraron un valor de R^2 de 0,9829 al enfrentar el número de granos de uva visibles en las imágenes con el determinado mediante etiquetado manual. En esa misma investigación, se encontró también una fuerte correlación entre el número de granos visibles detectados en las imágenes y el número absoluto de granos que contenían las plantas, incluyendo la fracción de ellos no visibles y, por tanto, no detectables en las imágenes. Ello resultó

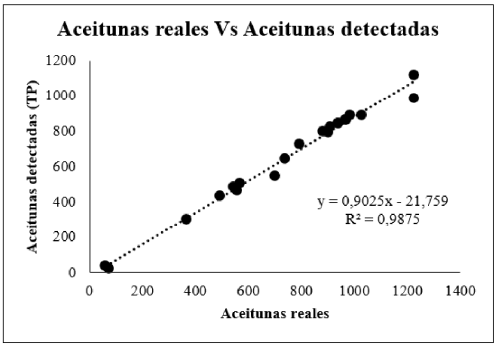


Figura 6: Resultado del estudio de correlación realizado enfrentando el número de aceitunas detectadas por el algoritmo con el número real.

en la emisión de una estimación de cosecha en gramos con un error medio del 1,3%. Por todo ello, el valor de R^2 obtenido en este trabajo argumenta a favor de que, si se encuentra una fuerte correlación entre el número de aceitunas detectadas mediante análisis de imagen y las totales contenidas en el árbol, se puede llegar a una metodología precisa de estimación temprana de la producción del olivar.

4 CONCLUSIONES

Este artículo presenta una metodología de análisis de imagen para la detección de aceitunas en imágenes de olivo, encaminada al desarrollo de un sistema de estimación de la producción del olivar. Dicha metodología implementa un preprocesamiento dirigido a hacer viable la detección de las aceitunas en las imágenes, bajo cualquier circunstancia de apariencia; mediante una red neuronal convolucional. Los resultados obtenidos muestran fuertes indicios de que el desarrollo presentado es un paso certero en la consecución del objetivo descrito. Además, es esperable que la metodología sea aplicable a otras variedades de aceituna gracias a que, en una fase temprana de maduración como la considerada en este trabajo, las características morfológicas y colorimétricas presentan una variabilidad contenida.

Como trabajo futuro, se pretende automatizar la toma de imágenes con el empleo de robots autónomos, así como mejorar las condiciones de captura con un enfoque de iluminación de mayor rendimiento. A su vez, se diseñarán experimentos de campo dirigidos a poder correlacionar la información extraída por análisis de imagen con valores reales de producción.

Agradecimientos

Esta investigación fue financiada por el Programa de Cooperación INTERREG V-A España-Portugal (POCTEP), y cofinanciada por fondos FEDER.

Los autores también quieren agradecer a la Cooperativa Virgen de la Oliva (Gibraleón, Huelva, Andalucía, España) por generosamente prestar sus olivares para el desarrollo de la experimentación de campo necesaria para culminar este trabajo.

English summary

OLIVE IDENTIFICATION AND COUNTING IN DIGITAL IMAGES TAKEN IN OLIVE ORCHARDS USING MATHEMATICAL MORPHOLOGY AND CONVOLUTIONAL NEURAL NETWORKS

Abstract

Early and accurate yield estimation is a very valued objective for modern agriculture. In the case of oliviculture, it is especially relevant due to the high economic value of its production. This paper presents a methodology aimed at achieving that end. Concretely, it comprises an artificial vision algorithm able to detect those olives that are visible in a digital image of an olive tree, captured directly in the field, at night-time and with artificial illumination. First, the image is preprocessed by means of mathematical morphology techniques and statistical filtering to, from this output, generate a subset of images with high probability of containing an olive. Thus, this preprocessing reduces the search space in a magnitude of 10^3. Next, these subimages are classified by a convolutional neural network as 'olive' or 'discarded'. From a total of 304,483 subimages, extracted from 21 images, the net correctly classified 98.23% of cases, and gave a coefficient of determination R^2 of 0.9875 when facing the number of detected olives to the real one. This achieved accuracy indicates that the found algorithm constitutes a solid step towards the implementation of a future system for early yield estimation of olive orchards.

Keywords: Precision agriculture, yield estimation, olive, artificial vision, convolutional neural network.

Referencias

[1] Food and Agriculture Organization of the United Nations (FAOSTAT). [Online]. Available: http://www.fao.org/faostat/en/#home.

[2] Saulle, R., La Torre, G., (2010) "The Mediterranean Diet, recognized by UNESCO as a cultural heritage of humanity", *Italian Journal of Public Health*, vol. 7(4), pp. 414-415.

[3] Rockström, J., Williams, J., Daily, G., Noble, A., Matthews, N., Gordon, L., Wetterstrand, H., DeClerck, F., Shah, M, Steduto, P., de Fraiture, C., Hatibu, N., Unver, O., Bird, J., Sibanda, L., Smith, J., (2010) "Sustainable intensification of agriculture for human prosperity and global sustainability", *Ambio*, vol. 46(1), pp. 4-17.

[4] Organization of United Nations (ONU). [Online]. Available: https://www.un.org/en/development/desa/news/population/2015-report.html

[5] Zarco-Tejada, P., Hubbard, N., Loudjani, P., (2014) "Precision Agriculture: An Opportunity for EU Farmers—Potential Support with the CAP 2014-2020". *Joint Research Centre (JRC) of the European Commission*. [Online]. Available: http://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL-AGRI_NT%282014%29529049

[6] Orlandi, F., Sgromo, C., Bonofiglio, T., Ruga, L., Romano, B., Fornaciari, M., (2010) "Yield modelling in a Mediterranean species utilizing cause-effect relationships between temperature forcing and biological processes". *Scientia Horticulturae*, vol. 123(3), pp. 412-417.

[7] Aguilera, F., Ruiz-Valenzuela, L. (2014) "Forecasting olive crop yields based on long-term aerobiological data series and bioclimatic conditions for the southern Iberian Peninsula". *Spanish Journal of Agricultural Research*, vol. 12(1), pp. 215-224.

[8] European Commission–Agriculture and Rural Development, (2011) "Agricultural Markets Brief–Brief 1: High commodity price and volatility: what lies behind the roller coaster ride?". [Online]. Available: http://ec.europa.eu/agriculture/analysis/tradepol/commodityprices/market-briefs/01_en.pdf.

[9] Nuske, S., Achar, S., Bates, T., Narasimhan, S., Singh, S., (2011) "Yield estimation in vineyards by visual grape detection". In *Proceedings of the International Conference on Intelligent Robots and Systems (IEEE/RSJ)*. San Fracisco, USA, Sept. 25-30, pp. 2352-2358.

[10] Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., Singh, S., (2014) "Automated visual yield estimation in vineyards". *Journal of Field Robotics*, vol. 31(5), pp. 837-860.

[11] Font, D., Tresanchez, M., Martínez, D., Moreno, J., Clotet, E., Palacín, J., (2015) "Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night". *Sensors*, vol. 15(4), pp. 8284-8301.

[12] Aquino, A., Millan, B., Diago, M.P., Tardaguila, J., (2018) "Automated early yield prediction in vineyards from on-the-go image acquisition". *Computers and electronics in agriculture*, vol. 144, pp. 26-36.

[13] Balasundram, S.K., Memarian, H., Khosla, R., (2013) "Estimating oil palm yields using vegetation indices derived from Quickbird". *Life Sci. J*, vol. 10(4), pp. 851-860.

[14] Chang, K.W., Shen, Y., Chung, L.J., (2005) "Predicting rice yield using canopy reflectance measured at booting stage". *Agronomy Journal* vol. 97, pp. 872-878.

[15] Fang, H., Liang, S., Hoogenboom, G., (2011) "Integration of MODIS LAI and vegetation index products with the CSM-CERES maize model for corn yield estimation". *International Journal of Remote Sensing*, vol. 32, pp. 1039-1065.

[16] Hayes, M.J., Decker, W.L., (1996) "Using NOAA AVHRR data to estimate maize production in the United States corn belt". *International Journal of Remote Sensing*, vol. 17, pp. 3189-3200.

[17] Bargoti, S., Underwood, J., (2017) "Deep fruit detection in orchards". In *Proceedings of the International Conference on Robotics and Automation (IEEE/ICRA)*, Singapore, Jun. 29-3, pp. 3626-3633.

[18] Stein, M., Bargoti, S., Underwood, J., (2016) "Image based mango fruit detection, localisation and yield estimation using multiple view geometry". *Sensors*, vol. 16(11), pp. 1915.

[19] Bargoti, S., Underwood, J.P., (2017) "Image segmentation for fruit detection and yield estimation in apple orchards". *Journal of Field Robotics*, vol. 34(6), 1039-1060.

[20] Fornaciari, M., Orlandi, F., Romano, B., (2005) "Yield forecasting for olive trees". *Agronomy Journal*, vol. 97(6), pp. 1537-1542.

[21] Galán, C., Vázquez, L., Garcia-Mozo, H., Dominguez, E., (2004) "Forecasting olive (Olea europaea) crop yield based on pollen emission". *Field Crops Research*, vol. 86(1), pp. 43-51.

[22] Minero, F.J.G., Candau, P., Morales, J., Tomas, C., (1998) "Forecasting olive crop production based on ten consecutive years of monitoring airborne pollen in Andalusia (southern Spain)". *Agriculture, Ecosystems & Environment*, vol. 69(3), pp. 201-215.

[23] Smith, W.J., (2007) Modern Optical Engineering, 4th ed. The Design of Optical Systems. McGraw-Hill Education – Europe, USA.

[24] Soille, P., (2004) Morphological Image Analysis. Principles and Applications, 2nd ed. Springer – Verlag, Berlin, Germany.

[25] Aquino, A., Diago, M.P., Millán, B., Tardáguila, J., (2017) "A new methodology for estimating the grapevine-berry number per cluster using image analysis". *Biosystems engineering*, vol. 156, pp. 80-95.

[26] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., (1998) "Gradient-based learning applied to document recognition". *Proceedings of the IEEE,* vol. 86(11), pp. 2278-2324.

[27] Krizhevsky, A., Sutskever, I., Hinton, G.E., (2012) "Imagenet classification with deep convolutional neural networks". In *Proceedings of the 25th international conference on neural information processing systems*, Lake Tahoe, Nevada, USA, Dec. 3-6, vol. 1, pp. 1097-1105.

[28] Simonyan, K., Zisserman, A., (2014) "Very deep convolutional networks for large-scale image recognition". *Cornell University arXiv*:1409.1556. [Online]. Available: https://arxiv.org/abs/1409.1556

[29] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., (2015) "Going deeper with convolutions". In *28th IEEE conference on computer vision and pattern recognition (CVPR)*, Boston, USA, pp. 1–9.

[30] He, K., Zhang, X., Ren, S., Sun, J., (2015) "Deep residual learning for image recognition". *Cornell University arXiv*:1512.03385. [Online]. Available: https://arxiv.org/abs/1512.03385

[31] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., (2017) "Inception-v4, inception-resnet and the impact of residual connections on learning". In *31st AAAI Conference on Artificial Intelligence (AAAI-17)*, San Francisco, California, USA, pp. 4278-4284.

[32] Qian, N, (1999) "On the momentum term in gradient descent learning algorithms". *Neural networks: the official journal of the International Neural Network Society*, 12(1), pp. 145–151.

4.3. Individual olive-tree characterisation from aerial imagery

4.3.1. Article 5

Automated Identification of Crop Tree Crowns from UAV multispectral Imagery by Means of Morphological Image Analysis

R. Sarabia, A. Aquino, J. M. Ponce, G. López, J. M. Andújar

Published in:

Journal: Remote Sensing (ISSN: 0168-1699)

Editorial: MDPI

Reference: vol. 12, no. 5, p. 748

Year: 2020

DOI: 10.3390/rs12050748

Quality index (Journal Citation Reports®, 2019): 9/30 (Q2) in the category "Remote Sensing". Impact Factor of 4.509.

remote sensing

MDPI

Article

Automated Identification of Crop Tree Crowns from UAV Multispectral Imagery by Means of Morphological Image Analysis

Ricardo Sarabia [1], Arturo Aquino [2], Juan Manuel Ponce [2,*], Gilberto López [1] and José Manuel Andújar [2]

[1] Instituto de Ingeniería Agrícola y Uso Integral del Agua, Chapingo Autonomous University, Texcoco 56230, Mexico; al17129576@chapingo.mx (R.S.); GLOPEZC@chapingo.mx (G.L.)
[2] Department of Electronic Engineering, Computer Systems and Automation, Higher Technical School of Engineering, University of Huelva, Fuerzas Armadas Ave, 21007 Huelva, Spain; arturo.aquino@diesia.uhu.es (A.A.); andujar@diesia.uhu.es (J.M.A.)
* Correspondence: jmponce.real@diesia.uhu.es

Received: 11 December 2019; Accepted: 21 February 2020; Published: 25 February 2020

Abstract: Within the context of precision agriculture, goods insurance, public subsidies, fire damage assessment, etc., accurate knowledge about the plant population in crops represents valuable information. In this regard, the use of Unmanned Aerial Vehicles (UAVs) has proliferated as an alternative to traditional plant counting methods, which are laborious, time demanding and prone to human error. Hence, a methodology for the automated detection, geolocation and counting of crop trees in intensive cultivation orchards from high resolution multispectral images, acquired by UAV-based aerial imaging, is proposed. After image acquisition, the captures are processed by means of photogrammetry to yield a 3D point cloud-based representation of the study plot. To exploit the elevation information contained in it and eventually identify the plants, the cloud is deterministically interpolated, and subsequently transformed into a greyscale image. This image is processed, by using mathematical morphology techniques, in such a way that the absolute height of the trees with respect to their local surroundings is exploited to segment the tree pixel-regions, by global statistical thresholding binarization. This approach makes the segmentation process robust against surfaces with elevation variations of any magnitude, or to possible distracting artefacts with heights lower than expected. Finally, the segmented image is analysed by means of an ad-hoc moment representation-based algorithm to estimate the location of the trees. The methodology was tested in an intensive olive orchard of 17.5 ha, with a population of 3919 trees. Because of the plot's plant density and tree spacing pattern, typical of intensive plantations, many occurrences of intra-row tree aggregations were observed, increasing the complexity of the scenario under study. Notwithstanding, it was achieved a precision of 99.92%, a sensibility of 99.67% and an F-score of 99.75%, thus correctly identifying and geolocating 3906 plants. The generated 3D point cloud reported root-mean square errors (RMSE) in the X, Y and Z directions of 0.73 m, 0.39 m and 1.20 m, respectively. These results support the viability and robustness of this methodology as a phenotyping solution for the automated plant counting and geolocation in olive orchards.

Keywords: aerial imagery; image analysis; multispectral imagery; crop tree; phenotyping; plant population; UAV

1. Introduction

Currently, global food demands entail one of the most challenging problems addressed by society. Indeed, as a consequence of the population growth expectations, the demand for crop production

is estimated to increase on the order of 100% in 2050, when compared to 2005 reports [1]. This scenario forces society to develop agricultural and food systems prone to proactively satisfy such a demand while being capable of minimizing the environmental impact. In this sense, crop phenotyping constitutes a crucial tool in order to achieve this balance.

Indeed, deep knowledge about observable crop trails and the way the genotype of plants expresses in relationship with the environmental factors comprise a relevant and valuable information for farmers [2]. Within this context, individual plant counting is a key factor, not only regarding to crop phenotyping, but also providing valuable information, supporting farmers when planning breeding strategies and another agricultural tasks. Thus, the plant population determines the crop density, defined as the number of plants per cultivated hectare. This statistic is closely related to different aspects, such as the efficiency of water and fertilizer resources, or pathogen susceptibility [3]. In addition, it plays a key role when estimating crop yield in tree-based cultivation, and it helps farmers when designing watering and/or fertilization schemes [4]. The importance of the plant population does not stop here, as it is a significant indicator when applying for public subsidies [5], pricing plantations [6], or assessing losses after any kind of extraordinary event, such as fire damage, pest infestations or other natural disasters. However, traditional counting methods are usually based on in-field human visual inspections, so as happens with other phenotyping activities [7,8], it implies tedious, time consuming and prone-to-error tasks, especially when it comes to large-scale plantations [3]. Due to these difficulties, there is a pressing need for the development of new techniques aimed at carrying out plant counting in an accurate, efficient and automated way.

Nowadays, Unmanned Aerial Vehicles (UAVs) have popularised as part of the remote sensing technologies incorporated into precision agriculture, and they have become widely used in crop phenotyping research [9,10]. This is mostly due to the advantages they offer over traditional aerial imaging systems already tested within this application, such as those based on manned airplanes or satellites. When compared to them, UAV-based imaging implies lower operational costs, less weather constraints and the possibility of operating under cloudy conditions [9,11–13]. Furthermore, the growth that the market related to UAVs and remote sensing equipment is experiencing nowadays makes this technology increasingly accessible and affordable. Hence, they are definitely promising tools within the scope of smart farming and precision agriculture, with potential uses in crop phenotyping tasks [9,14].

In fact, when focusing on plant detection and counting, a considerable amount of research where crop tree identification is realised from UAV-based imagery can be found already. Images acquired are usually processed, generating representative data structures of the study sites which are subsequently analysed in order to detect and count the plants. Hence, Malek et al. [5] approached palm tree detection, by analysing a set of candidates, previously computed using the scale-invariant feature transform (SIFT), with an extreme learning machine (ELM) classifier. Candidates categorised as trees were post-processed by means of a contour method based on level sets (LS) and local binary patters (LBP), in order to identify the shapes of their crowns. In Miserque-Castillo et al. [15], a framework for counting oil palms was developed, where a sliding window-based technique procured a set of candidates. After processing with LBP, they were classified by a logistic regression model. Primicerio et al. [16] studied plant detection within vine rows. The segmentation of the plant mass was carried out on the basis of dynamic segmentation, Hough space clustering and total least squares regression. After individual plant identifications were estimated, a multi-logistic model for the detection of missing plants was applied. Jiang et al. [17] introduced a GPU-accelerated scale-space filtering methodology for detecting papaya and lemon trees in UAV images. To that end, initial captures were converted to a Lab-based colour space, mostly exploiting the information contained in the channel a (representative of the colour values from red to green) to differentiate the plants from the ground. Koc-San et al. [18] undertook citrus trees location and counting from UAV multispectral imagery. To that end, they proposed a set of procedures based on sequential thresholding and the Hough transform. In the same vein, Csillik et al. [19] focused on citrus crops, intending the identification of trees by using convolutional neural networks (CNNs). In addition, they used a simple linear iterative clustering (SLIC)

algorithm for classification refinement. CNNs were also used by Ampatzidis and Partel [20] in order to detect citrus trees. Specifically, the CNN model was trained by using a YOLOv3 object detection algorithm. Furthermore, they implemented a normalised difference vegetation index (NDVI)-based image segmentation method for estimating the canopy area. In Selim et al. [21], approached orange tree detection from high-resolution images, by applying an object-based classification methodology, using a multi-resolution segmentation of the data derived from aerial imagery. Deep learning and CNN technology was exploited by Aparna et al. [4]. In this case, coconut palm tree detection was the aim. Initial captures were transformed into an HSV colour representation, and then binarized and conveniently cropped in sub-images, with which the CNN classifier was trained. In Kestur et al. [22], an ELM methodology was proposed for detecting tree crowns from aerial images captured in the visible spectrum. Thus, the developed ELM spectral classifier was applied in order to segment the tree crowns-pixel areas from the rest of the image. The methodology was validated by studying banana, mango and coconut palm trees. Marques et al. [23] focused on the detection of chestnut trees. They considered different kinds of sensorics for acquiring aerial images. Thus, RGB and Colour Infrared (CIR) images were used in their research, where different segmentation techniques were explored in order to properly isolate the tree-belonging pixel-regions to subsequently carry out the eventual identification of the trees.

Regarding olive plantations, which constitute the study case considered throughout the experimentation developed here, several studies where olive tree phenotyping is approached by using UAV-based aerial imagery can be found. Thus, Díaz-Varela et al. [24] attempted the estimation of the height and crown diameter of olive trees by means of structure-from-motion (SfM) image reconstruction and geographical object-based image analysis (GEOBIA). Along the same line, Torres-Sánchez et al. [25] also proposed a methodology for the estimation of different olive tree features. Particularly, height, crown volume and canopy area were addressed. This was accomplished by generating digital surface models (DSMs) from aerial imagery, and object-based image analysis (OBIA). This study was extended in [26], where different flight altitudes and overlapping degrees were tested in order to optimise the DSM generation, in terms of computational cost. In Salamí et al. [6], olive trees counting was approached by using a UAV equipped with a small embedded computer. This device was aimed at processing captures on board, and to provide via cloud services, nearly real-time plant count estimations to the end-user.

In this paper, a new methodology for the identification of crop trees located in intensive farming-based orchards, by means of the analysis of aerial images, is proposed. To that end, we start from a set of aerial captures acquired by a UAV equipped with a multispectral camera while flying over the land plot under study. These multispectral images are processed in order to yield a DSM, following standard image matching and photogrammetry techniques. The core of the novel proposal of the methodology is comprised by an image analysis-based algorithm, aimed at identifying the trees by exploiting the elevation information contained in this data structure. To that end, the DSM is converted into as a greyscale image, where elevation information is approached as grey level values. Then, this image is transformed by means of mathematical morphology, in order to individually segment the tree-belonging pixels from the ground, by a statistical global thresholding-based binarization. Eventually, that resulting segmentation is analysed by an ad-hoc procedure to detect intra-row tree aggregations, consisting in studying the second central moment of the tree pixel-regions. The whole methodology was tested in an intensive olive orchard, obtaining results that highlight its effectiveness as a full-automated solution for crop trees detection and counting, and its robustness against complex scenarios, as intra-row tree aggregations and a strong ground elevation variability were present in the study plot.

Hereafter, the present manuscript is structured as follows: Section 2 focuses on the experimental design. Thus, Section 2.1 describes the characteristics of the olive orchard in which, as study case, images were acquired for the purpose of testing the methodology. Section 2.2 exposes all the aspects related to how aerial image acquisition was performed. In Section 2.3, the image analysis methodology for

trees detection, counting and geolocation is developed, addressing the stages of image pre-processing (Section 2.3.1), the generation of a DSM as a base data structure (Section 2.3.2), and the image segmentation and analysis (Sections 2.3.3 and 2.3.4, respectively). Then, in Section 2.4, the set of metrics computed to assess the performance of the methodology is proposed. Section 3 presents the results obtained, which are then discussed in Section 4. Section 5 concludes the manuscript, giving a brief summary of the main findings achieved and identifying aspects that might be approached in further investigations. Finally, Appendix A formally defines all the morphological operators used throughout the developed image analysis methodology.

2. Materials and Methods

2.1. Study Case Site

The olive grove where the testing aerial imagery was acquired is located in Gibraleón, province of Huelva (Andalusia, Southwest Spain). In particular, the area under study, centred in the coordinates 7°02'48.44''W and 37°20'39.80''N, corresponds to an orchard with an approximate extent of 17.5 ha, where an intensive cultivation system is applied, with a plant spacing pattern of 5.5 × 7 m; the *Olea europaea* L. cultivated variety is Picual. It should be noted that this orchard shows a notable variability in terms of soil composition, crown size of the trees and altitude, varying from around 54 m to around 96 m above sea level. A third-party aerial capture of the study site, obtained by manned flight-based imaging, is shown in Figure 1. It should be underscored that this third-party image is only offered for the purpose of illustrating the study plot, so it was not used at all throughout the experimentation.

Figure 1. Third-party aerial capture of the case study site shown to illustrate the study plot, highlighted in red.

2.2. Image Acquisition

2.2.1. Aerial Imaging Equipment

Aerial imaging was conducted using a DJI™ Matrice 100 UAV (SZ DJI™ Technology Co., Ltd., Shenzhen, Guangdong, China). This device is propelled by four rotors (quadcopter), enabling its vertical take-off and landing. With a diagonal wheelbase of 650 mm and a maximum take-off weight of 3600 g, it can reach a maximum cruise speed of 22 m/s, withstanding a wind resistance up to 10 m/s. It is controlled in an operating frequency varying from 5725 to 5825 GHz, with a maximum transmission distance of 5 km.

Images were taken with the multispectral camera MicaSense RedEdge-M™ (MicaSense, Inc., Seattle, WA, USA), installed on the UAV. This sensing device is capable of capturing information in five different spectral bands within the visible and the infrared spectrum. Table 1 summarises the most relevant features related to these bands.

Table 1. Features of the spectral bands captured by the multispectral camera MicaSense RedEdge-M™.

Band Number	Band Name	Centre Wavelength (nm)	Bandwidth (nm)
1	Blue	475	20
2	Green	560	20
3	Red	668	10
4	Near Infrared	840	40
5	Red Edge	717	10

The camera was mounted together with a dedicated GPS device for the purpose of georeferencing each captured image. A downwelling light sensor (DLS) was also included into the setup, in order to calibrate the images according to the changing conditions of ambient light. Finally, for accurate ground reflectance calibration, a reference board (grey reference) was used by imaging it during both the take-off and landing. In Figure 2, the UAV is shown together with all the equipment described above.

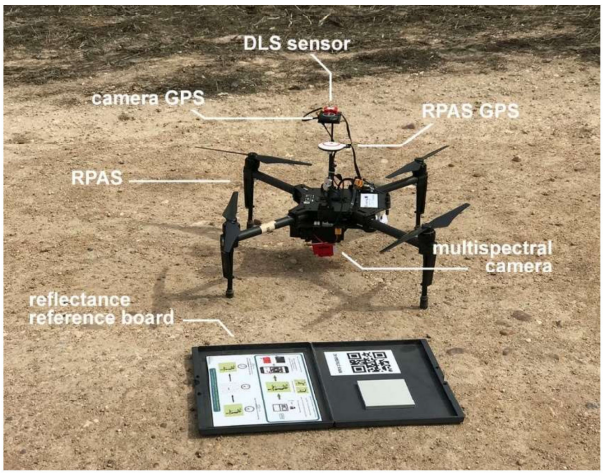


Figure 2. Equipment used to capture the aerial imagery used in this paper.

2.2.2. Flight Planning and Development

The flight mission planning was set with the DJI™ Flight Planner software, by drawing the polygon delimiting the study plot (highlighted in red in Figure 1). Within the study plot, the mission was planned according to the criterion of minimising the number of turns to be made by the UAV to cover it entirely. Thus, the flight was configured to be performed autonomously, at an altitude of 70 m and at a cruise speed of 15 km/h. The multispectral camera was configured with a time period between captures of 1.5 s. With these settings, it was intended to capture images with forward and lateral overlaps of 85% and 65% respectively, and with a desired GSD of 0.05 m/pixel. The flight took place on June 13, 2019, approximately between 11 a.m. and 1 p.m. Litchi software (VC Technology, Ltd. ©, London, UK) was used for operating and monitoring the mission. A total of 44,325 images were acquired during the flight, 8865 per each of the five spectral bands in which the multispectral camera can capture information.

2.3. Image Analysis Methodology for Olive Trees Detection, Geolocation and Counting

The main objective pursued in this investigation is the development of a procedure able to perform olive tree detection, location and counting from aerial captures by means of image analysis. To that end, a methodology has been designed under those principles to, first, transform images acquired into a DSM, as a representative data structure of the whole orchard under study; and then, to exploit the information contained in it in order to carry out a binary segmentation, in which tree-belonging

pixels could be differentiated from the rest of the image. Eventually, the result of this segmentation is analysed to detect intra-row aggregations, thus finally yielding the individual tree locations and the accurate plant population estimation. The flowchart shown in Figure 3 illustrates the different stages comprising the developed methodology, which are deeply detailed throughout the next subsections.

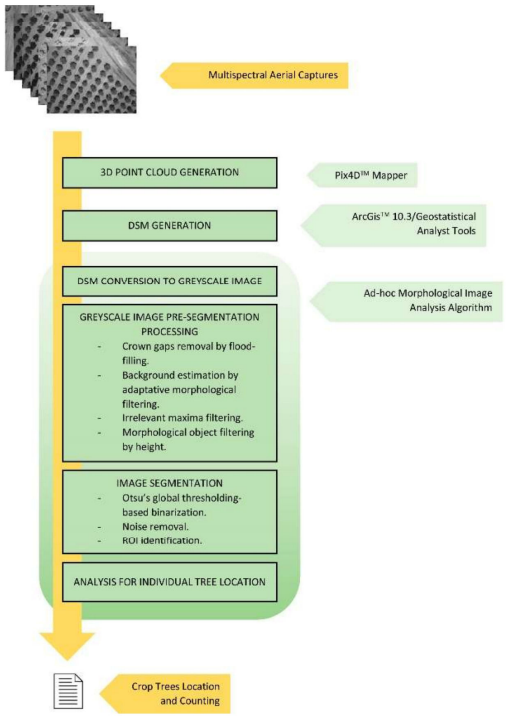


Figure 3. Representative diagram of the methodology proposed for detecting and counting crop trees from multispectral aerial images.

For simplicity purposes, all morphological operators involved in the methodology described throughout this section, are formally defined in Appendix A.

2.3.1. Image Pre-Processing

As a first step, captures obtained by aerial imaging are radiometrically corrected using the illumination information provided by the camera's DLS sensor and the reflectance measured in the images captured of the reference board. Then, the corrected images are processed to yield the orthomosaics corresponding to each of the five spectral bands considered. In Figure 4, a colour image resulting from the combination of the blue, near infrared (NIR) and red edge bands is presented. It should be underlined that this ad-hoc image was exclusively generated for the purpose of supporting the assessment of the methodology's performance by a human observer, as detailed in Section 2.4. Therefore, they were chosen so as to obtain a proper visual tree differentiation, being other combinations of bands surely also suitable for this purpose.

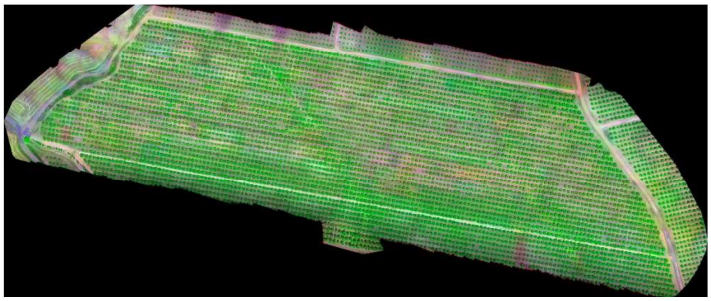


Figure 4. Colour image generated from the information provided by the orthomosaics of the Blue, Red Edge and NIR spectral bands.

In addition, a 3D point cloud is generated as well, to later develop the representative DSM of the overflown land plot. To that end, every point in the cloud is determined by its re-projection in at least three images; then, it is characterised with a triplet of coordinates, where the two first ones determine its relative location within the cloud and the third one refers to its elevation. Thus, a high-density 3D point cloud with a total number of 205,998,922 points is reached.

The task of creating both the set of orthomosaics and the 3D point cloud was carried out using the photogrammetry Pix4D^{TM} Mapper software. As representative indicators of the errors committed during the pre-processing stage, the software reported root-mean square errors (RMSE) in the X, Y and Z directions of 0.73 m, 0.39 m and 1.20 m, respectively. Note that these errors do not correspond to the quality of the point cloud, but to the error between the initial and the computed image positions.

2.3.2. Digital Surface Model (DSM)

The DSM is generated by deterministic spatial analysis, from the 3D point cloud yielded above, by applying an inverse distance weighting (IDW) interpolation [27]. According to this method, the attribute value (the elevation in this case) of an unsampled point is decided from the attribute value of its surrounding known points. The influence of the known sampled points decreases as their distance from the targeted point increase, so the unsampled point value is computed on the basis of the attribute values of the surrounding points observations, inversely weighted according to their distance. So, being S_0 a targeted point, its interpolation value $\hat{Z}(S_0)$ can be mathematically defined as follows:

$$\hat{Z}(S_0) = \sum_{i=1}^{N} \lambda_i Z(S_i). \quad (1)$$

where $Z(S_i)$ is the observed value for the i-th surrounding point S_i of N points; λ_i is the weight assigned to S_i, according to its distance d_{i0} to S_0. Hence, λ_i can be defined as follow:

$$\lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^{N} d_{i0}^{-p}} \quad (2)$$

where p is a weighting exponent that controls the way in which weight decreases with distance; the weights λ_i vary between 0 and 1 for each point and the total sum of them is the unit: $\sum_{i=1}^{N} \lambda_i = 1$.

For computing the DSM using this IDW spatial interpolation, ArcGis^{TM} 10.3 (Esri, Inc., Redlands, CA, USA) and its Geostatistical Analyst Tools extension were used. The size of the cell was matched to the cell size of the orthomosaics computed before. In the same vain, interpolation output raster was also restricted by the dimensions of these orthomosaics. In addition, it should be noted that, during the analysis, it was stablished a fixed neighbourhood search, using a circular radius distance of 10 m and a maximum number of neighbourhood points of 4, with a weighting exponent p of 2.

2.3.3. Image Segmentation Algorithm

The DSM, obtained after processing the initial aerial captures, is used as fundamental data to eventually perform crop trees detection and the subsequent location and counting. Every voxel (3D pixel) in the DSM is defined by its x and y position within the map, and its altitude with respect to the sea level. This altitude information is exploited in such a way that trees are segmented by considering their absolute height with respect to their local neighbourhood.

First, the DSM is approached as a 2D greyscale image by taking the voxels' elevation information as the intensity values of their corresponding pixels in this greyscale image. Thus, given DSM as the representative matrix of the DMS previously computed, the intensity matrix, GS_{DSM}, which approaches this model as an 8-bit greyscale image, can be defined as follows:

$$GS_{DMS}(x,y) = \begin{cases} DSM(x,y), \text{ if } DSM(x,y) > 0 \\ 0, \text{ in any other case} \end{cases}. \quad (3)$$

where $DSM(x,y)$ is the elevation value with respect to the sea level provided by the DSM for the point (x,y). Figure 5 shows a representation of the DSM as a greyscale image.

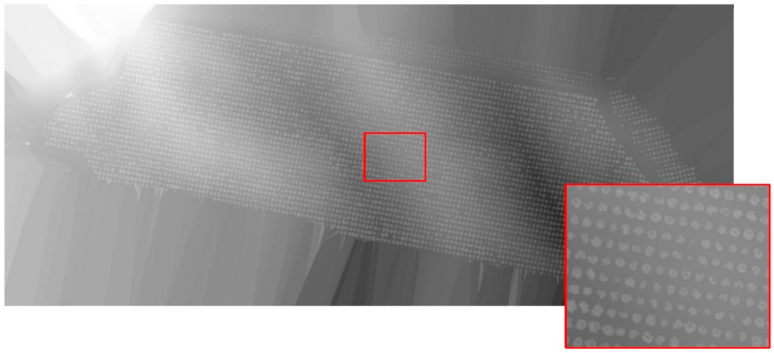
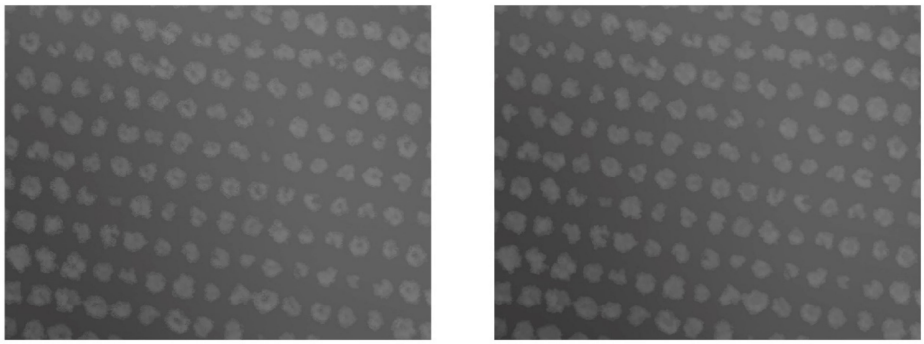


Figure 5. Representation of the computed DSM as the intensity image GS_{DSM}. Note in the zoomed area, highlighted in the red square, the differences in terms of grey level between those pixel regions which apparently belong to olive trees, and those from the surrounding ground. Then, given that each pixel intensity value is assigned according to its elevation in the DSM, higher pixel values indicate higher altitudes with respect to the sea level. It should be noted that, for the sake of facilitating its visualisation, the image display range has been established between the minimum bigger-than-zero value from the DMS, and its maximum.

Once this greyscale image is obtained, a filling operation is performed to homogenise the grey level values of the tree crowns which, in some cases, showed darker areas potentially related to hollows in the foliage. Mathematically, it can be defined from a morphological reconstruction as follows:

$$\begin{aligned} I_{GS1} &= R^{\varepsilon}_{GS_{DMS}}(GS'_{DMS}), \\ GS'_{DMS}(x,y) &= \begin{cases} GS_{DMS}, & \text{if } (x,y) \text{ is a border pixel} \\ 255, & \text{in any other case} \end{cases}. \end{aligned} \quad (4)$$

where GS'_{DMS} is a border image of GS_{DMS}, and R^{ε} refers to the morphological reconstruction by erosion (ε) of GS_{DMS} from marker GS'_{DMS} until idempotence. Figure 6 shows the effect of this operation on the zoomed area of Figure 5.



(**a**) (**b**)

Figure 6. Filling gaps illustration: (**a**) same zoomed area of GS_{DMS} shown in Figure 5; (**b**) result of the filling gaps operation applied to (**a**).

Afterwards, a homogenisation of the grey level values of I_{GS1} is performed aimed at favouring its later optimum binarization. Since I_{GS1} directly derives from the DSM, its pixel values represent altitude magnitudes expressed in meters with respect to the sea level. Consequently, disturbing cases when binarizing may appear, such as that in which ground pixels have higher grey level values than those of tree pixels (when the former are at higher altitudes than the latter). Hence, in order to avoid this difficulty, I_{GS1} is homogenised by subtracting from it an accurate background estimate. This is calculated by iteratively opening I_{GS1} with a circular structuring element of increasing radius, taking at each step the minimum value between the opening results at the i-th and the i-1-th iteration. Mathematically:

$$I_{BE_{DEF}} = I_{BE_n},$$

being

$$I_{BE_i} = \begin{cases} I_{GS1}, & \text{if } i = 0 \\ MIN\left(I_{BE_{i-1}}, \gamma_{\beta_i}\left(I_{BE_{i-1}}\right)\right) & \text{in any other case} \end{cases}, \quad i = 1, \ldots, n. \tag{5}$$

where γ_{β_i} is the morphological opening operation using a disc-shaped structuring element β of radius $i \times 5$. For a given tree crown in the image, its optimum filtering takes place when its grey level values are substituted with the minimum value existing in its closest background neighbourhood. It happens when the opening operation is performed using a structuring element with the minimum radius allowing the element to completely contain the tree crown. Therefore, note that the formulated approach provides a flexible framework favouring the accurate filtering of every tree independently from its size. The number of iterations has been fixed to $n = 14$, which corresponded to a maximum radius value of the structuring element equal to 70. This value has been set to ensure the accurate filtering of the greater trees, being adaptable to different image capturing conditions deriving in other maximum tree crown sizes. Once $I_{BE_{DEF}}$ is computed, the homogenisation of I_{GS1} is obtained by:

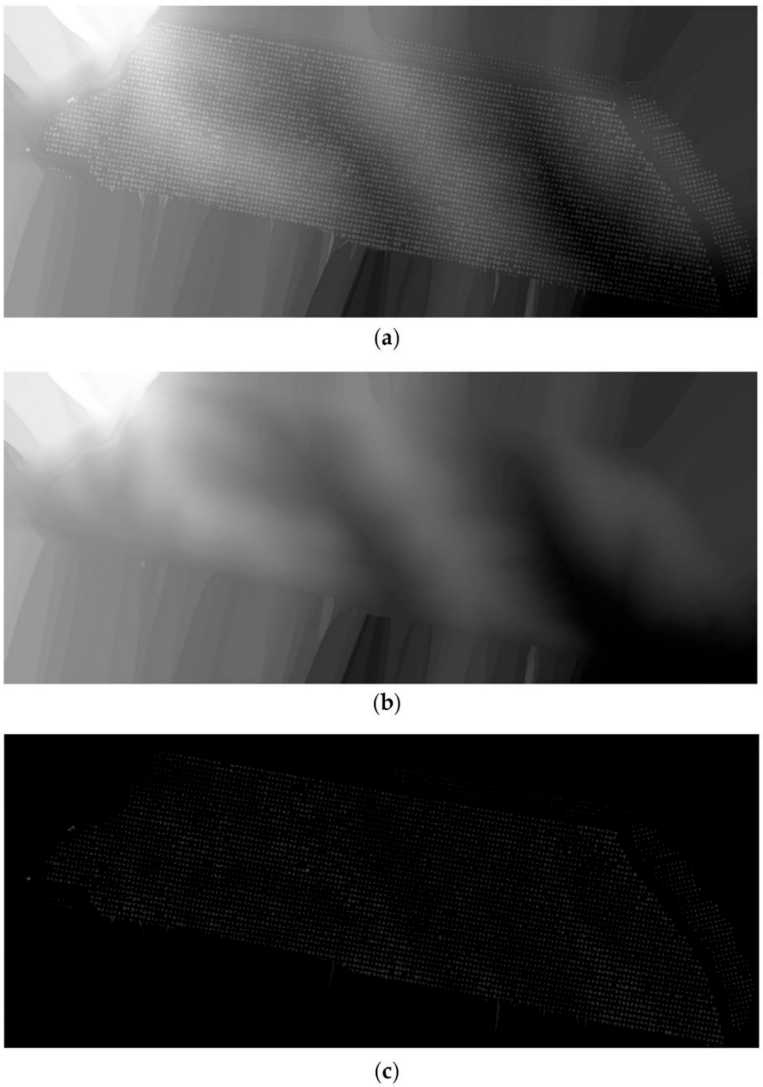
$$I_{GS2} = I_{GS1} - I_{BE_{DEF}}. \tag{6}$$

Figure 7 illustrates the described process to yield a homogenised version of I_{GS1}. At this point, I_{GS2} is a homogenised image in which its grey level values represent absolute altitudes. In other words, the previous processing made the effect of ideally flattening the original surface and placing its background at the sea level. In this context, tree crowns are the elements expected to be at higher levels, so the next processing is intended for removing from the image irrelevant maxima, which were

considered to be those pixels with altitude values lower or equal than 1 m. This effect is achieved by applying the H-maxima transform to image I_{GS2}:

$$I_{GS3} = HMAX_h(I_{GS2}) = R^{\delta}_{I_{GS2}}(I_{GS2} - h),\ h = 1. \quad (7)$$

where R^{δ} refers to the morphological reconstruction by dilation (δ) of I_{GS2} from marker $I_{GS2} - h$. In this case, artefacts with elevation values greater than 1 m were retained for being of interest; note that this criterion can be easily modified by adjusting the h parameter.



(**a**)

(**b**)

(**c**)

Figure 7. (**a**) Greyscale image I_{GS1} resulting from filling gaps in the image shown in Figure 5, GS_{DMS}; (**b**) background estimation of (**a**), $I_{BE_{DEF}}$; (**c**) resulting image I_{GS2} after subtracting (**b**) to (**a**).

Next, the elements surviving the previous filtering by height are segmented by binarizing image I_{GS3} using the Otsu's method [28]. This approach assumes that the population of grey level values of the image is made up of two dominant groups or classes, corresponding to the foreground and the

background pixels, respectively. Hence, it determines the grey value maximising the separability of both classes, which results in the greater median distance between them, or analogously, the minimum intra-class variance. Therefore, given the threshold *thresh* resulting from applying the Otsu's method to image I_{GS3}, its binarization can be defined as follows:

$$I_{BIN1}(x,y) - \begin{cases} 255, \text{ if } I_{GS3}(x,y) > \text{thresh} \\ 0, \text{ in any other case} \end{cases}. \quad (8)$$

Figure 8 illustrates I_{BIN1}, resulting from the binarization of image I_{GS3}, shown in Figure 7c.

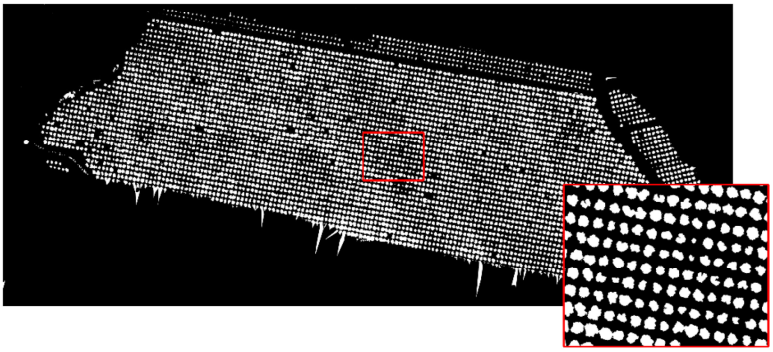


Figure 8. Image I_{BIN1} resulting from the binarization of I_{GS3}, shown in Figure 6c. Note in the zoomed area, in the red square, how potential plants have been accurately segmented from the background.

As a result of this binarization, image pixels are segmented into two classes, background (black pixels) and foreground (white pixels); this latter being potentially formed by pixels belonging to olive trees. Next, in order to remove the spurious connected components (set of neighbour foreground pixels) abnormally small, a morphological opening is applied on the binary image I_{BIN1}:

$$I_{BIN2} = \gamma_\beta(I_{BIN1}). \quad (9)$$

where γ_β stands for the morphological opening performed by using a disk-shaped structuring element β of 5 pixels in radius. To exactly recover the shape of the connected components surviving to this noise filtering, I_{BIN1} is morphologically reconstructed by dilation (R^δ) from marker I_{BIN2}, which leads to I_{BIN3}:

$$I_{BIN3} = R^\delta_{I_{BIN1}}(I_{BIN2}). \quad (10)$$

After noise removal, the polygon drawn to delimit the region of interest (ROI) for flight planning (Section 2.2.2) is used as a mask image, I_{ROI}, to constrain the area of interest within the image for the rest of the analysis. Figure 9 illustrates I_{ROI}, together with the result of its application to I_{BIN3}, which can be mathematically formulated as:

$$I_{BIN\text{def}}(x,y) = \begin{cases} I_{BIN3}(x,y), \text{ if } I_{ROI}(x,y) > 0 \\ 0 \text{ in any other case} \end{cases}. \quad (11)$$

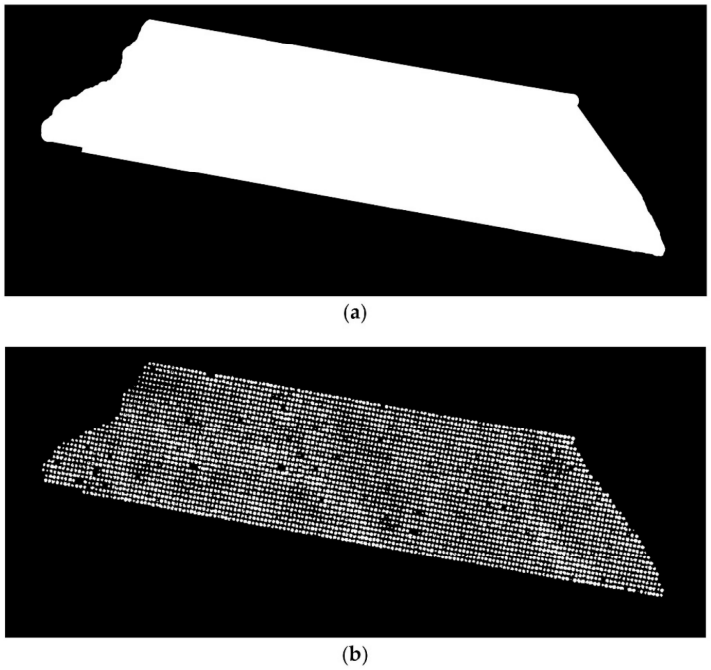


Figure 9. (**a**) ROI mask image I_{ROI}; (**b**) image I_{BINdef} resulting from filtering the binary image I_{BIN3} (visually very similar to I_{BIN3}, shown in Figure 8), with image I_{ROI}.

2.3.4. Image Analysis Algorithm for Tree Counting and for the Estimation of Tree Locations

As it can be seen in Figure 9, I_{BINdef} provides a segmentation of the olive trees from the background. A first approach to count the number of plants might just consider the number of connected components in that binary image. Nevertheless, the possibility of existing connected components not exactly corresponding to a sole olive tree has to be considered. Indeed, because of the variability in terms of crown size shown by the trees of the plot under study, their foliage may appear overlapped within the same row, thus resulting in wrongly merged connected components in the binary image; overlapping of tree crowns from different rows is not expected as it is prevented by pruning; Figure 10 illustrates this phenomenon. Therefore, the image analysis procedure described below has been developed intended to accurately provide plant population, despite intra-row tree aggregations.

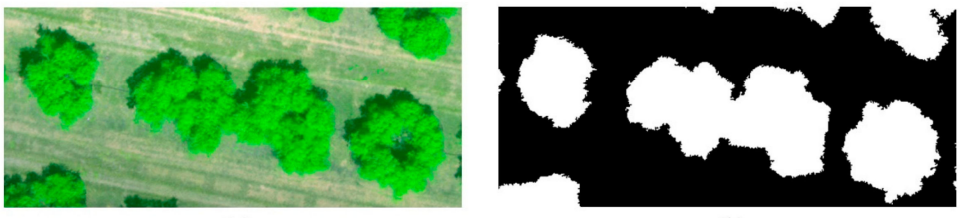


Figure 10. (**a**) Sub-image of the study plot orthomosaic represented in Figure 4, where it can be observed a couple of trees with overlapping foliage; (**b**) sub-image of the binary image resulted from the segmentation performed, I_{BINdef}, corresponding to the area represented in (**a**). Note how the two olive trees share the same connected component.

The procedure is based on analysing the morphology of the segmented connected components of I_{BINdef}, in order to determine the estimated number of trees contained in them. To that end, the components of the binary image are firstly approached with the ellipses that share the same normalised second central moment [29]. Thus, for a given connected component cc_i, its representing ellipse E_i is defined by the following set of elements:

$$E_i = \left\{c_xE_i, c_yE_i, d_1E_i,\ d_2E_i, \alpha E_i\right\}. \tag{12}$$

being c_xE_i and c_yE_i the coordinates of its centre, d_1E_i and d_2E_i the length in pixels of its two axes and αE_i the angle formed by its longer axis and a horizontal imaginary axis. Consequently, the length of the major and minor axes of the ellipse can be defined as:

$$\begin{array}{c} MajAx(cc_i) = MAX\{d_1E_i, d_2E_i\}, \\ MinAx(cc_i) = MIN\{d_1E_i, d_2E_i\}. \end{array} \tag{13}$$

As can be seen in Figure 11, whilst the minor axes keep comparable length values throughout the whole population of ellipses, regardless of the number of plants contained in the corresponding components, the length of the major axes show a strong dependence with this number. To exploit this feature, next, the maximum length value of all the computed minor axes was calculated as a reference to subsequently be used throughout the rest of the analysis. Thus:

$$MAX_{MinAX} = MAX\{MinAx(cc_i)\}. \tag{14}$$

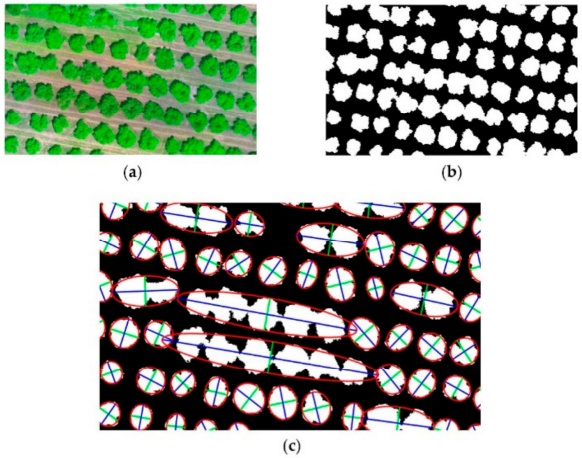


Figure 11. (**a**) Sub-image of the study plot orthomosaic represented in Figure 4; (**b**) sub-image of the binary image resulted from the segmentation performed, corresponding to the area represented in (**a**); (**c**) representation of the ellipses (in red) computed for each connected component in the image (**b**), with their corresponding major (in blue) and minor (in green) axes.

Then, counting of trees was conducted by comparing MAX_{MinAX} to the length of the major axis of each connected component cc_i, by computing:

$$TreeNumber(cc_i) = \begin{cases} 1, \text{ if } MajAx(cc_i) \leq MAX_{MinAX} \times 1.20 \\ [\![MajAx(cc_i)/MAX_{MinAX}]\!], \text{ in any other case} \end{cases}. \tag{15}$$

As the shapes of tree crowns are irregular, the MAX_{MinAX} value, computed on the population of minor axes, might be slightly lower than the length of the major axis of an eventual connected

component representing a sole tree. Additionally, we note that tree spacing in agricultural plantations, such as that considered in this study, is ideally regular, so the potential for overlapping trees are greater. Therefore, as Equation (15) shows, increasing MAX_{MinAX} by 20% provides flexibility to the former situation while it respects the latter assumption, as the aggregation of great tree crowns is not expected to enlarge the resulting object by only a 20%. The concrete value has been decided empirically, being not critical, as values moderately higher and lower were also found to provide comparable results. Finally, once the number of plants per connected component is estimated, the total number of trees is calculated by the addition of these partial results:

$$TotalTreePopulation = \sum_{i=1}^{n} TreeNumber(cc_i), \tag{16}$$

being n the number of connected components.

Finally, once trees are counted, a representative location for each of them within the image is attempted. Hence, the following definition was established:

$$TreeLocation_{cc_i} = \left\{ \left(x_{ij}, y_{ij}\right) \middle| \begin{array}{c} (x_{i1}, y_{i1}) = \left(c_xE_i, c_yE_i\right),\ TreeNumber(cc_i) = 1 \vee \\ (x_{ik}, y_{ik}),\ TreeNumber(cc_i) = t,\ k = 1, \ldots, t \end{array} \right\}. \tag{17}$$

This is, for a given connected component cc_i containing a sole tree, the location of the latter is decided as the location of the centre of the ellipse E_i representing the former. Conversely, for aggregated components, the location of the contained multiples trees is estimated by equally spacing them throughout the major axis of its representing ellipse, taking as reference the centre of this last. With this approach, two situations must be considered. The first case refers to when cc_i contains an odd number of trees. In this scenario, the location of the central tree matches with the centre of its ellipse, being the resting locations estimated by displacements to the left and to the right of this reference. Mathematically:

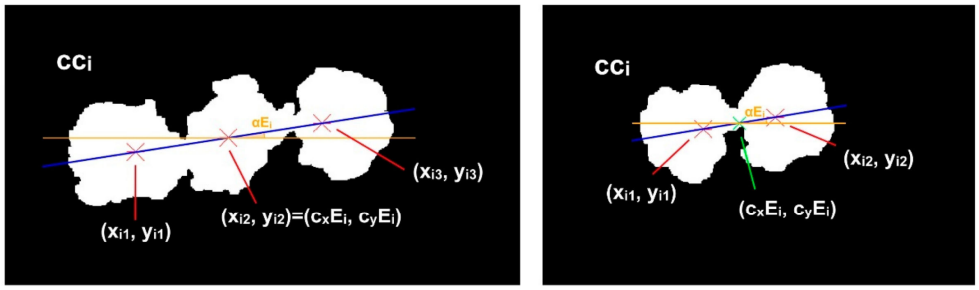
$$\begin{array}{c} x_{ik} = \left\{ \begin{array}{c} c_xE_i + k \times jump_i \times \cos(\alpha E_i + \pi),\ if\ k < t/2 \\ c_xE_i,\ \ if\ k = t/2 \\ c_xE_i + (k - t/2) \times jump_i \times \cos(\alpha E_i),\ if\ k > t/2 \end{array} \right., \\ y_{ik} = \left\{ \begin{array}{c} c_yE_i + k \times jump_i \times \sin(\alpha E_i + \pi),\ if\ k < t/2 \\ c_yE_i,\ \ \ if\ k = t/2 \\ c_yE_i + (k - t/2) \times jump_i \times \sin(\alpha E_i),\ if\ k > t/2 \end{array} \right., \\ k = 1, \ldots, t,\ \ t = TreeNumber(cc_i), \\ jump_i = MajAx(cc_i)/(t+1). \end{array} \tag{18}$$

where $jump_i$ represents the magnitude of the displacements among the estimated tree centres. Note that the first case models the estimated locations placed at the left of the central tree, the third case models those placed at the right, and the second one defines the case of the central tree. The second scenario occurs when cc_i contains an even number of trees. For this case, the centre of its representing ellipse does not match with the expected centre of a tree, but with the overlapping zone of two of them. Hence, this location is not assigned to any tree, but it is only taken as a reference:

$$\begin{array}{c} x_{ik} = \left\{ \begin{array}{c} c_xE_i + (k - 0.5) \times jump_i \times \cos(\alpha E_i + \pi),\ if\ k \leq t/2 \\ c_xE_i + (k - t/2 - 0.5) \times jump_i \times \cos(\alpha E_i),\ if\ k > t/2 \end{array} \right., \\ y_{ik} \left\{ \begin{array}{c} c_yE_i + (k - 0.5) \times jump_i \times \sin(\alpha E_i + \pi),\ if\ k \leq t/2 \\ c_yE_i + (k - t/2 - 0.5) \times jump_i \times \sin(\alpha E_i),\ if\ k > t/2 \end{array} \right., \\ k = 1, \ldots, t,\ \ t = TreeNumber(cc_i), \\ jump_i = MajAx(cc_i)/(t+1). \end{array} \tag{19}$$

The first case models the estimated locations at the left of the centre of the ellipse E_i, while the second case models those placed at its right. Figure 12 graphically describes the formulated procedure

to estimate tree locations. In Figure 13, the result of computing the tree potential location points is illustrated. The yielded locations are marked in red in the binary sub-image shown in Figure 11.



(**a**) (**b**)

Figure 12. Illustration of the process to estimate a representative location for aggregated trees, formulated in Equations (17)–(19). Examples for odd (**a**) and even (**b**) number of aggregated trees are given.

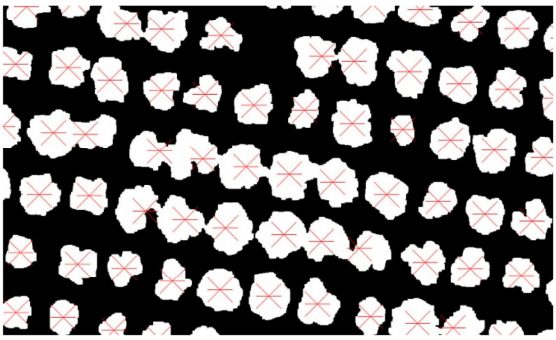


Figure 13. Result of the estimation of the individual tree location points.

2.4. Performance Evaluation of the Image Analysis Methodology

In order to assess the performance of the methodology proposed, it was firstly necessary to locate and determine the exact number of olive trees in the land plot under study. This was carried out by a human observer, by inspecting, labelling and counting the tree crowns appearing in the ad-hoc orthomosaic of the study site proposed in Figure 4.

The performance assessment of the methodology was approached by comparing the actual number of plants, and their distribution, to the estimations yielded by the image analysis algorithm.

In order to quantitatively evaluate this comparison, the set of metrics defined here below are proposed:

- *Precision*: it gives the hit ratio for the trees found by the algorithm. Mathematically:

$$Precision = \frac{TP}{TP + FP}. \tag{20}$$

where TP (true positives) is the number of trees correctly identified, and conversely, FP (false positives) refers to the number of instances wrongly proposed by the algorithm as potential olive trees. A tree is considered to be correctly identified only when the algorithm placed its estimated location within its crown.

- *Sensitivity*: it provides the ratio of actual trees found by the algorithm:

$$Sensitivity = \frac{TP}{TP + FN}. \quad (21)$$

where *FN* (false negatives) is defined as the number of actual olive trees not detected by the algorithm.

- $F1_{score}$: it is the harmonic mean of the two metrics described above, being mathematically defined as:

$$F1_{score} = 2 \times \frac{(Precision \times Sensitivity)}{(Precision + Sensitivity)} \quad (22)$$

3. Results

According to the metrics proposed, the results provided by the presented methodology for crop trees detection, location and counting are exposed in Table 2. As it can be observed, 99.92% of tree proposals were correct, and 99.67% of the actual trees were found.

Table 2. Performance assessment of the automated trees detection and counting methodology, expressed in terms of the metrics defined to that purpose.

Actual Tree Population	Estimated Tree Population	*TP*	*FP*	*FN*	*Precision*	*Sensitivity*	$F1_{score}$
3919	3909	3906	3	13	0.9992	0.9967	0.9975

Regarding the failures detected, and focusing on the false positives (*FP*) reported, each of them can be justified by a different reason. Thus, one of them was caused by a car that was parked very close to the study site. Because of its height, it could not be discarded during image processing, neither filtered when the image was cropped according to the specified region of interest. As a result, a very small residual connected component, corresponding to this vehicle, was inevitably considered when analysing the ultimate binary image. A second false positive resulted from a tree with an anomalously damaged crown, so it was detected by the algorithm as two different plants. Finally, a last false tree detection was obtained when processing a large connected component containing seven aggregated olive trees. Due to the morphology and disposition of the overlapped tree pixel regions, the number of plants contained was overestimated in one unit. The different issues related to the false positives detected during the assessment of the methodology are illustrated in Figure 14.

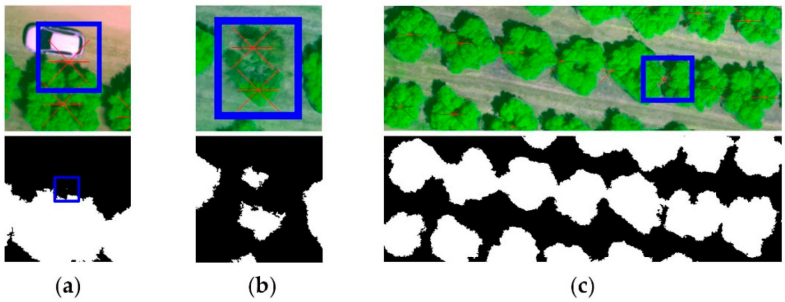
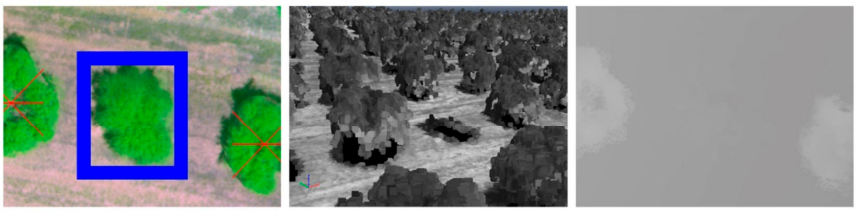


Figure 14. False positives detected during performance assessment: (**a**) case related to a car parked next to the study site; (**b**) case consequence of wrongly splitting one tree into two different connected components because of its damaged condition; (**c**) case obtained after overestimating the number of trees contained in an aggregated connected component.

With respect to false negatives (*FN*), one of them was detected to come from the absence of information in the DSM, this probably due to not having enough matching points from different captures when reconstructing this part of the image. As a result, the elevation information in those corresponding points, provided by the DSM, was not significant enough to enable the discrimination of this plant (the phenomenon is illustrated in Figure 15).



(**a**) (**b**) (**c**)

Figure 15. False negative resulting from a lack of information in the point cloud: (**a**) aerial sub-image where the tree wrongly discarded by the algorithm is represented; (**b**) 3D point cloud-based representation of the area shown in (**a**); (**c**) elevation information provided by the DSM, represented as a greyscale image, corresponding to the area shown in (**a**).

In this respect, it should be noted that the density and quality of the 3D point cloud used to generate the DSM, is directly related to the overlapping with which aerial imagery is captured [25]. As commented in Section 2.2.2, the image acquisition flight and the multispectral camera setup were planned for the purpose of achieving a forward overlap of 85%. By increasing this overlapping, results could be virtually improved. However, since 99.97% of the trees were properly reconstructed, i.e., 3918 among 3919, it seems plausible to consider the setup proposed for image capture as valid. Being discarded defects in flight and image capture parametrisation, it is difficult to determine the reasons that provoked this issue, but it might be probably related to problems when capturing the aerial images, either due to weather conditions that could occasionally affect the stability of the UAV, or due to problems with the operability of the camera. Meanwhile, the rest of false negative cases detected were related to small trees, most of them in growth stage, which did not reach the minimum height (1 m) to be properly segmented from the background.

4. Discussion

Table 3 compares the results of the methodology presented in this paper to those of the main published works also aimed at automated crop tree detection in orchards. A first aspect to be highlighted is that the present work outperforms the other proposals, despite the fact it was tested on a considerably greater plant population when compared to most of the reported research. Consequently, this surely included a wider variability in terms of the individual characteristics of the trees, and the way they are disposed throughout the land plot under study. Also, it should be underscored that, contrary to most of those works, this study considered challenging conditions related to overlapping intra-row tree crowns, aspect with a special impact on the accuracy with which plant population can be estimated in intensive orchards.

Table 3. Comparison of performance of different methods for crop trees counting published in the bibliography, and the present work.

Method	Actual Tree Population	*Precision*	*Sensitivity*	$F1_{score}$
Torres-Sánchez et al., 2015 [25]	135	–	0.945–0.969	–
Torres-Sánchez et al., 2018 [26]	–	–	0.970	–
Salamí et al., 2019 [6]	332	0.9939	0.9909	0.9924
Malek et al. [5]	617	0.9009	0.9440	0.9219
Csillik et al. [19]	2912	0.9459	0.9794	0.9624
Ampatzidis and Partel [20]	4931	0.9990	0.9970	0.9980
Selim et al. [21]	105	–	0.8286	–
Marques et al. [23]	1092	0.9944	0.9780	0.9861
This work	3918	0.9992	0.9967	0.9975

Thus, focusing on the case of the olive, a crop around which the proposed methodology was validated, counting of trees based on aerial imagery was attempted in Salamí et al. [6], obtaining a remarkable average precision of 99.84%. Nevertheless, plant detection was approached by using a circular template, imposing the prerequisite of only considering isolated trees, thus preventing that their crowns could appear overlapped in aerial captures. Contrary, the methodology presented here was able to deal with the individual location and counting of 385 trees configuring 293 aggregated connected components. Only in the case shown in Figure 14, the number of trees contained in such a component was not properly estimated. Moreover, the replicability of the methodology presented in [6] is questionable, as trees segmentation was attempted by colour discrimination. Indeed, it is very probable that any kind of natural or artificial artefact with similar colour to that of the olive tree canopies, could generate false positives. In this case, the precision of the colour segmentation approach, and consequently of the subsequent trees detection and counting, is compromised. A segmentation also based in pixel reflectance, although not only in the visible bands, followed by OBIA analysis was used in Torres-Sánchez et al. [25]. Concretely, a multi-resolution segmentation was firstly performed using the DSM and the green and NIR bands, considering colour, shape, smoothness and compactness, for which threshold values were manually adjusted. The manual decision of such key segmentation parameters questions concerns about its replicability in different situations. Furthermore, the approach requires a subsequent OBIA analysis to filter the first segmentation results. Conversely, the methodology described here proposes an analytical solution to the segmentation problem, only making use of the h parameter (equation (7)) in the segmentation step. In addition, this is a comprehensible parameter as it represents the minimum desired height in meters for the trees to be segmented. Then, h is more likely to be seen as a configuration parameter rather than a performance one. On a set of 135 olive trees, the study presented in Torres-Sánchez et al. [25] yielded sensitivity values ranging from 0.945 to 0.969, not considering the case of overlapping tree crowns. Later, the same main author and others assessed the influence of image overlap in the quality of the resulting DSM [26]. The methodology described in [25] was slightly modified and tested on an indeterminate number of trees, corroborating in the best scenario of those tested a sensibility of 0.97 in olive trees counting. The case of overlapping trees was not faced either.

Beyond the olive case, in Malek et al. [5], an overall precision of 0.9009 when detecting palm trees was achieved. They proposed a method based on training an ELM classifier on a set of key points, potentially representative of the occurrences of the trees, extracted from the initial captures. Csillik et al. [19] made also use of machine learning, concretely CNNs, for detecting citrus trees. Ampatzidis and Partel [20] also focused their research on citrus orchards, and also using CNN-based tree location. Despite the fact all these studies reported solid results, it should be noted that these kinds of machine learning solution tend to be strongly linked to the visual features of the crown trees with which they are trained. This fact makes their direct application to different kinds of crops difficult, but it surely implies the generation of new training sets and models. Contrary, the method proposed in this paper comprises an analytical solution, based on the morphological analysis and characterisation

of the general features of trees within the frame of an intensive cultivation, thus not being linked to a concrete type of crop. In Selim et al. [21], it was proposed a method for detecting orange trees from aerial imagery. The problem was undertaken in this case by means of object-based image analysis, correctly detecting 87 out of the 105 trees visible in the orthomosaic of the study case. Nevertheless, as with other researches previously referenced, difficulties were reported when dealing with overlapping tree crowns. In Kestur et al. [22], tree detection was faced on the basis of ELM- spectral and spatial classification. Despite promising results were reported about identifying trees belonging to different crops, it was not clearly specified how the training set was generated, hindering the replicability of the methodology. Marques et al. [23] proposed an effective method for detecting chestnut trees, clustering the plants by exploiting elevation data and vegetation indices (VI) information. About this latter, it should be noted that VI-based segmentations are strongly dependent of the spectral reflectivity features of the vegetation cover present in the study sites. Indeed, depending of its nature, the coverage may be confused with the plants aimed to be identified, thus potentially increasing the number of false positives yielded. This phenomenon may affect the generality of the proposed solution.

5. Conclusions

This investigation was undertaken in order to design and evaluate a framework for the automated identification, geolocation and counting of crop trees in intensive cultivation areas by means of UAV-based aerial imagery, multispectral sensing and image analysis techniques. The results reported support the viability of the methodology proposed as a valuable tool in phenotyping tasks, within the scope of the precision agriculture.

After testing in an olive orchard with 3919 trees, 99.67% of the plants were rightly identified, outperforming the results given by previous published work. Indeed, the algorithm designed for segmenting and analysing the data structure obtained from aerial captures, based on morphological image processing principles and the statistical analysis of the moments of tree-corresponding pixel artefacts, showed a remarkable performance in terms of tree discrimination, achieving very high detection rates. In addition, the solution also showed to be solid when dealing with multiple intra-row overlapping tree crowns. These findings should also be framed within the context of the complexity of the considered scenario, since the study plot was outstandingly larger than those used in most of previous studies, and it presented a remarkable variability in terms of soil composition, elevation and amount of weed.

Future work will test the application of the presented methodology to other types of orchards. In addition, it would be interesting to assess the performance of the algorithms when dealing with different plant spacing patterns, all of this for the sake of increasing confidence in the generality of the proposed solution.

Author Contributions: Conceptualization, R.S., A.A. and G.L.; methodology, R.S., A.A. and J.M.P.; software, A.A.; validation, R.S., A.A. and J.M.P.; formal analysis, R.S.; investigation, R.S., A.A. and J.M.P.; resources, R.S.; data curation, R.S., A.A. and J.M.P.; writing—original draft preparation, R.S.; writing—review and editing, A.A., J.M.P. and J.M.A.; visualization, R.S. and A.A.; supervision, A.A., J.M.A. and G.L.; project administration, J.M.A.; funding acquisition, J.M.A and G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research and APC were funded by the INTERREG Cooperation Program V-A SPAIN-PORTUGAL (POCTEP) 2014–2020, and co-financed with ERDF funds, grant number 0155_TECNOLIVO_6_E, within the scope of the TecnOlivo Project.

Acknowledgments: Authors would like to thank "Virgen de la Oliva" olive-oil cooperative for generously offering their orchards to conduct this work. R.S. would also like to thank Mexican National Council of Science and Technology (CONACYT) for supporting the development of this investigation.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Mathematical morphology is a non-linear image processing technique, built from the basis of the set theory, essentially aimed at analysing the relevant structures in the image by proving this with a set

called structuring element (SE), which has an a-priori known shape and size. This appendix briefly defines the morphological operators employed in this paper, suggesting the reader to consult [30,31] for a deeper study.

Let f be a greyscale image representing a mapping from a subset D_f of $\mathbb{Z}^2$, which defines de domain of the image, into a bounded subset of nonnegative integers N_0:

$$f : D_f \subset \mathbb{Z}^2 \rightarrow N_0 = \{0, \ldots, t_{max}\} \subset \mathbb{Z}. \tag{A1}$$

where t_{max} is the maximum value allowing to reach the data type used (e.g., 1 for binary images, 256 for 8-bit images, etc.). Thus, f maps the correspondence element by element between two sets, the first being composed of spatially ordered elements ρ (pixels), $\rho \in D_f$ and denoted by a pair of coordinates (x, y), while the second is built with an ordered set of possible values.

With the previous definitions, the intersection of two greyscale images, f and g, is defined as:

$$f(\rho) \wedge g(\rho) = min[f(\rho), g(\rho)]. \tag{A2}$$

where *min* refers to the minimum operation. Conversely, the union of those two images responds to the following equation:

$$f(\rho) \vee g(\rho) = max[f(\rho), g(\rho)], \tag{A3}$$

being *max* the maximum operation.

The SE is an essential tool in mathematical morphology, used to study the shape of the objects contained in an image. Mathematically, an SE element can be seen as a binary image β, defining a mapping of a subset D_β of $\mathbb{Z}^2$ to the subset of integer binary values B_0:

$$\beta : D_\beta \subset \mathbb{Z}^2 \rightarrow B_0 = \{0, 1\} \subset \mathbb{Z}. \tag{A4}$$

With this definition, β maps the correspondence between the spatially ordered pixels ρ, $\rho \in D_\beta$ and referenced by a pair of coordinates (x, y), and their values. This mapping must be designed so as to morphologically describe the object to be analysed, being necessary for its application that $\#\big(D_\beta\big) < \#\big(D_f\big)$. Common shapes implemented with SEs include circles, lines, diamonds, etc. In practice, the SE is used as a kernel, with its origin in its central pixel. Hence, an image is proven pixel by pixel to this kernel, modifying at every step the pixel in the image matching with the central pixel of the kernel, according to a given operation.

The morphological erosion of image f by an SE β, this last being centred in pixel ρ, is given by the expression:

$$\Big[\varepsilon_\beta(f)\Big](\rho) = min\Big\{f(\rho + b)\Big|b \subset D_\beta\Big\}. \tag{A5}$$

Therefore, pixel ρ in image f is modified with the minimum value of its neighbourhood according to the filter implemented by SE β. The effect of erosion is the expansion of darker regions, conditioned by the shape defined in SE.

The dual operator of erosion is dilation. The morphological dilation of image f by a SE β centred in pixel ρ, is formulated as:

$$\Big[\delta_\beta(f)\Big](\rho) = max\Big\{f(\rho + b)\Big|b \in D_\beta\Big\}. \tag{A6}$$

By duality, dilation expands brighter regions in f according to the morphology of SE.

Combining erosion and dilation, two new operators called opening (γ) and closing (φ) are obtained:

$$\gamma_\beta(f) = \delta_\beta\Big(\varepsilon_\beta(f)\Big), \tag{A7}$$

$$\varphi_\beta(f) = \varepsilon_\beta\Big(\delta_\beta(f)\Big). \tag{A8}$$

Opening removes those brighter objects in the image that can be completely covered by β. Dually, closing removes the darker objects in the image completely covered by the SE.

The operators described are complemented by geodesic transformations. The geodesic dilation is the iterative dilation of an image f, called marker, using a unitary SE, with respect to the mask g. Marker f must be contained within mask g. Mathematically, the operator is defined as:

$$\begin{array}{l} \delta_g^{(n)}(f) = \delta_g^{(1)}\left[\delta_g^{(n-1)}(f)\right], \text{ being } \delta_g^{(1)}(f) = \delta(f) \wedge g, \\ \text{where :} \\ \#\left(D_f\right) = \#\left(D_g\right), \text{ and } f(\rho) \leq g(\rho), \ \forall \rho \in D_f, D_g. \end{array} \tag{A9}$$

Based on (9), the geodesic erosion of marker f constrained by mask g is:

$$\begin{array}{l} \varepsilon_g^{(n)}(f) = \varepsilon_g^{(1)}\left[\varepsilon_g^{(n-1)}(f)\right], \text{ being } \varepsilon_g^{(1)}(f) = \varepsilon(f) \vee g, \\ \text{where :} \\ \#\left(D_f\right) = \#\left(D_g\right), \text{ and } f(\rho) \geq g(\rho), \ \forall \rho \in D_f, D_g. \end{array} \tag{A10}$$

Geodesic dilation and erosion are the basis for building morphological reconstructions. Indeed, the morphological reconstruction by dilation of mask g by marker f, is the geodesic dilation of f constrained by g until idempotence. It is denoted by:

$$\begin{array}{l} R_g^{\delta}(f) = \delta_g^{(i)}(f), \\ \text{where : } i \text{ is such that :} \\ \delta_g^{(i)}(f) = \delta_g^{(i+1)}(f). \end{array} \tag{A11}$$

Consequently, the dual morphological reconstruction by erosion of mask g by marker f, is the geodesic erosion of f constrained by g until idempotence:

$$\begin{array}{l} R_g^{\varepsilon}(f) = \varepsilon_g^{(i)}(f), \\ \text{where : } \ i \text{ is such that :} \\ \varepsilon_g^{(i)}(f) = \varepsilon_g^{(i+1)}(f). \end{array} \tag{A12}$$

References

1. Tilman, D.; Balzer, C.; Hill, J.; Befort, B.L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 20260–20264. [CrossRef] [PubMed]
2. Araus, J.L.; Cairns, J.E. Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* **2014**, *19*, 52–61. [CrossRef] [PubMed]
3. Jin, X.; Liu, S.; Baret, F.; Hemerlé, M.; Comar, A. Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sens. Environ.* **2017**, *198*, 105–114. [CrossRef]
4. Aparna, P.; Ramachandra, H.; Harshita, M.P.; Harshitha, S.; Nandkishore, K.; Vinod, P.V. CNN Based Technique for Automatic Tree Counting Using Very High Resolution Data. In Proceedings of the 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, India, 25–28 April 2018; IEEE: Bangalore, India, 2018; pp. 127–129.
5. Malek, S.; Bazi, Y.; Alajlan, N.; AlHichri, H.; Melgani, F. Efficient Framework for Palm Tree Detection in UAV Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4692–4703. [CrossRef]
6. Salamí, E.; Gallardo, A.; Skorobogatov, G.; Barrado, C. On-the-Fly Olive Trees Counting Using a UAS and Cloud Services. *Remote Sens.* **2019**, *11*, 316. [CrossRef]
7. Shakoor, N.; Lee, S.; Mockler, T.C. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Curr. Opin. Plant Biol.* **2017**, *38*, 184–192. [CrossRef]
8. Furbank, R.T.; Tester, M. Phenomics – Technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* **2011**, *16*, 635–644. [CrossRef]

9. Sankaran, S.; Khot, L.R.; Espinoza, C.Z.; Jarolmasjed, S.; Sathuvalli, V.R.; Vandemark, G.J.; Miklas, P.N.; Carter, A.H.; Pumphrey, M.O.; Knowles, N.R.; et al. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *Eur. J. Agron.* **2015**, *70*, 112–123. [CrossRef]
10. Yang, G.; Liu, J.; Zhao, C.; Li, Z.; Huang, Y.; Yu, H.; Xu, B.; Yang, X.; Zhu, D.; Zhang, X.; et al. Unmanned Aerial Vehicle Remote Sensing for Field-Based Crop Phenotyping: Current Status and Perspectives. *Front. Plant Sci.* **2017**, *8*, 1111. [CrossRef]
11. Tripicchio, P.; Satler, M.; Dabisias, G.; Ruffaldi, E.; Avizzano, C.A. Towards Smart Farming and Sustainable Agriculture with Drones. In Proceedings of the 2015 International Conference on Intelligent Environments, Prague, Czech Republic, 15–17 July 2015; IEEE: Prague, Czech Republic, 2015; pp. 140–143.
12. Hunt, E.R.; Daughtry, C.S.T. What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture? *Int. J. Remote Sens.* **2018**, *39*, 5345–5376. [CrossRef]
13. Peña, J.M.; Torres-Sánchez, J.; de Castro, A.I.; Kelly, M.; López-Granados, F. Weed Mapping in Early-Season Maize Fields Using Object-Based Analysis of Unmanned Aerial Vehicle (UAV) Images. *PLoS ONE* **2013**, *8*, e77151. [CrossRef]
14. Floreano, D.; Wood, R.J. Science, technology and the future of small autonomous drones. *Nature* **2015**, *521*, 460–466. [CrossRef] [PubMed]
15. Miserque Castillo, J.Z.; Laverde Diaz, R.; Rueda Guzmán, C.L. Development of an aerial counting system in oil palm plantations. *IOP Conf. Ser. Mater. Sci. Eng.* **2016**, *138*, 012007. [CrossRef]
16. Primicerio, J.; Caruso, G.; Comba, L.; Crisci, A.; Gay, P.; Guidoni, S.; Genesio, L.; Ricauda Aimonino, D.; Vaccari, F.P. Individual plant definition and missing plant characterization in vineyards from high-resolution UAV imagery. *Eur. J. Remote Sens.* **2017**, *50*, 179–186. [CrossRef]
17. Jiang, H.; Chen, S.; Li, D.; Wang, C.; Yang, J. Papaya Tree Detection with UAV Images Using a GPU-Accelerated Scale-Space Filtering Method. *Remote Sens.* **2017**, *9*, 721. [CrossRef]
18. Koc-San, D.; Selim, S.; Aslan, N.; San, B.T. Automatic citrus tree extraction from UAV images and digital surface models using circular Hough transform. *Comput. Electron. Agric.* **2018**, *150*, 289–301. [CrossRef]
19. Csillik, O.; Cherbini, J.; Johnson, R.; Lyons, A.; Kelly, M. Identification of Citrus Trees from Unmanned Aerial Vehicle Imagery Using Convolutional Neural Networks. *Drones* **2018**, *2*, 39. [CrossRef]
20. Ampatzidis, Y.; Partel, V. UAV-Based High Throughput Phenotyping in Citrus Utilizing Multispectral Imaging and Artificial Intelligence. *Remote Sens.* **2019**, *11*, 410. [CrossRef]
21. Selim, S.; Sonmez, N.K.; Coslu, M.; Onur, I. Semi-automatic Tree Detection from Images of Unmanned Aerial Vehicle Using Object-Based Image Analysis Method. *J. Indian Soc. Remote Sens.* **2019**, *47*, 193–200. [CrossRef]
22. Kestur, R.; Angural, A.; Bashir, B.; Omkar, S.N.; Anand, G.; Meenavathi, M.B. Tree Crown Detection, Delineation and Counting in UAV Remote Sensed Images: A Neural Network Based Spectral–Spatial Method. *J. Indian Soc. Remote Sens.* **2018**, *46*, 991–1004. [CrossRef]
23. Marques, P.; Pádua, L.; Adão, T.; Hruška, J.; Peres, E.; Sousa, A.; Sousa, J.J. UAV-Based Automatic Detection and Monitoring of Chestnut Trees. *Remote Sens.* **2019**, *11*, 855. [CrossRef]
24. Díaz-Varela, R.; de la Rosa, R.; León, L.; Zarco-Tejada, P. High-Resolution Airborne UAV Imagery to Assess Olive Tree Crown Parameters Using 3D Photo Reconstruction: Application in Breeding Trials. *Remote Sens.* **2015**, *7*, 4213–4232. [CrossRef]
25. Torres-Sánchez, J.; López-Granados, F.; Serrano, N.; Arquero, O.; Peña, J.M. High-Throughput 3-D Monitoring of Agricultural-Tree Plantations with Unmanned Aerial Vehicle (UAV) Technology. *PLoS ONE* **2015**, *10*, e0130479. [CrossRef] [PubMed]
26. Torres-Sánchez, J.; López-Granados, F.; Borra-Serrano, I.; Peña, J.M. Assessing UAV-collected image overlap influence on computation time and digital surface model accuracy in olive orchards. *Precis. Agric.* **2018**, *19*, 115–133. [CrossRef]
27. Shepard, D. A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 23rd ACM National Conference, New York, NY, USA, 27–29 August 1968; ACM Press: New York, NY, USA, 1968; pp. 517–524.
28. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]
29. Jain, A.K. *Fundamentals of Digital Image Processing*; Prentice Hall: Englewood Cliffs, NJ, USA, 1989; ISBN 0133361659.

30. Soille, P. *Morphological Image Analysis: Principles and Applications*; Springer-Verlag GmbH: Heidelberg, Germany, 2004; ISBN 9783662050880.
31. Serra, J. *Image Analysis and Mathematical Morphology, vol. I*; Academic Press Inc.: Cambridge, MA, USA, 1982; ISBN 9780126372427.

4.3.2. Article 6

A Methodology for the Automated Delineation of Crop Tree Crowns from UAV-based Aerial Imagery by means of Morphological Image Analysis

J. M. Ponce, A. Aquino, D. Tejada, J. M. Andújar

(As of October 2020, in peer review process)

Chapter 5
Conclusions

Chapter 5. Conclusions

The most relevant findings and conclusions that can be drawn from the investigation conducted around the different topics addressed in this Thesis are summarised herebelow.

Thus, research conducted for characterisation of individual olive fruits, aimed at developing Machine Vision systems for automated grading, resulted highly productive. Regarding the image acquisition framework proposed, it should be remarked that the ad-hoc imaging chamber resulted very effective for the purpose of capturing high contrast images in terms of luminosity between the fruits and the background. In addition, the LED-based illumination system made it possible to avoid the appearance of any potential shadow cast by the fruits when photographed. This facilitated the initial background-foreground segmentation of the captures and significantly contributed to streamlining their processing, thus reducing the overall computational cost. Given that the proposed methodology was developed with the ultimate aim of being integrable into actual systems for automated fruit-grading, which should operate in real time, computational complexity was a critical issue to bear in mind.

Regarding feature estimation models, initial assumption of characterising the size of each fruit from the pixel-dimensions of the ellipse with the same normalized second central moment than the corresponding connected component resulted a smart approach, attending to external validation results. Indeed, variety-dependent models never exceeded relative errors of 0.80% and 1.05% when respectively estimating the major and the minor diameters of the fruits. The same conclusion can be drawn for mass estimation models, especially after introducing a pixel-weighing approach for mimicking the sphericity of the fruits from their plain binary representations, thus enhancing individual mass characterisation and consequently providing accurate estimates, with relative errors always below 1.16%. Nevertheless, with regard to these latter estimators, the better performance of variety-dependent models versus independent ones should be highlighted. In this sense, the higher error obtained after testing the variety-independent models can be explained by differences in physical characteristics related to the mass the fruits, as fruit density, stone-fresh ratio, or fat content, each of which may differ from one variety to another. In short, external validation results suggest that mass-size relationship may vary among olive varieties. This could enable a new line of investigation, so further research may explore non-linear modelling approaches in order to express this dependency when developing variety-independent mass estimation models.

Automated categorisation of olive-fruits according to variety was another challenge faced at this stage of research. In this vein, this issue was approached as an image classification problem, proposing a methodology based on the training of six different CNN implementations, and aimed at discriminating between seven varieties of olive-fruit. Image classifiers obtained reached remarkable results, hitting an overall top accuracy of 95.91% in the case of the Inception-Resnet V2 CNN implementation. In addition, this

developed technology is fully compatible with those mass-and-size predictive models above-mentioned. Hence, the integration of both methodologies into a single system opens up the possibility of developing a Machine Vision-device that could be installed into the conveyor belts where olive fruits are transported, with the capacity of carrying out the counting, and the categorisation of each of them according to its mass, size and variety.

Another topic of interest addressed in the Thesis has been the assessment of potential solutions, based on Computer Vision techniques, that would allow early yield estimation in an automatic and non-invasive way. To that end, the development of a procedure for detecting visible olive fruits in digital images covering entire olive-trees, directly taken in the field at night-time, and with artificial illumination, was undertaken. This issue was again approached as an image classification problem; in this case, by categorising a set of sub-images obtained during preprocessing according to whether they contain a fruit or not.

Different CNN architectures were considered and tested for that purpose. In this case, Inception-ResNetV2 outperformed other alternatives and resulted the most balanced topology, correctly classifying during external validation the 83.13% (*sensitivity, SE*) of those subimages/candidates which actually contained an olive-fruit; with a precision (*PR*) of 84.80%, and yielding a measured accuracy and $F1_{score}$ of 0.9822 and 0.8396, respectively. In any case, results reported support the validity of this approach as the core from which to developed a comprehensive system for image analysis-based early yield estimation, consisting of an autonomous vehicle that would automatically perform data acquisition, photographing olives from a lateral view, thus providing those images in which would be carried out automated fruit detection. In order to accomplish this objective in further research, some issues may be addressed. In this sense, and for the purpose of improving the overall performance of the algorithm proposed, it should be increased the size of the training set, with the inclusion of new samples rather than by using data augmentation techniques. On the other hand, the methodology could benefit from a new lighting system for image acquisition, which could provide a more homogeneous canopy illumination, facilitating the detection of fruits located in the peripheral zones of the trees.

In addition, the development of the OLIVEnet dataset, publicly available to the scientific community for supporting further research in this subject, should be highlighted as an important contribution of this work.

On the other hand, the last line of argument considered in the Thesis is the one related to the use of UAV-based aerial imagery in conjunction to image analysis techniques for developing mapping solutions for monitoring, inventorying and phenotyping crop-trees. The corresponding research carried out resulted in a comprehensive framework for the automated detection and geolocation of crop-trees, from the processing of multispectral images remotely sensed by using a UAV. Additionally, individual segmentation of each tree-crown present in the aerial representations was achieved, thus providing information about individual canopies, and enabling the possibility of estimating morphological characteristics of them.

It should be noted that the initial idea was to develop this technology to be applicable in olive orchards, but potentially extensible to other tree-based cultivars too. Because of that, it was intended tree identification by exploiting elevation information from 3D-based representations of the orchards, instead of using vegetation indices or any type of Deep Learning approach that could strengthen the dependence between the models achieved and the case studies from which they were developed.

Results obtained after testing the methodology in different cultivars reinforce the validity of the above-mentioned initial assumption. Indeed, the methodology performance was assessed in three different orchards, in each of which is farmed a different type of crop. In all case studies, detection rates resulted highly positives, and individual tree-canopies segmentation very accurate. Even in the orange grove case, in which it could be found numerous occurrences of overlapping between the canopies of adjacent trees, performance was rather remarkable, achieving a *precision* of 99.87%, and a *recall* of 97.82% when individually identifying the trees in the land plot surveyed. Despite these promising results, further investigation could address the study of new cultivars, with the aim of increasing the confidence in the generality of the solution proposed here.

References

1. Vossen, P. *Olive oil: History, production, and characteristics of the world's classic oils*; American Society for Horticultural Science, 2007; Vol. 42; ISBN 0018-5345.
2. Kashiwagi, K.; Yamna, E.; Arfa, L.; Zaibet, L. Growing Olive Oil Export and Intra-Industry Trade in Mediterranean Countries: Application of Gravity Model. *Sustainability* **2020**, *12*, 7027, doi:10.3390/su12177027.
3. International Olive Council (IOC) Word Olive Oil Figures Available online: https://www.internationaloliveoil.org/wp-content/uploads/2020/04/HO-W901-29-11-2019-P.pdf (accessed on Sep 24, 2020).
4. International Olive Council (IOC) World Table Olives Figures Available online: https://www.internationaloliveoil.org/wp-content/uploads/2020/04/OT-W901-29-11-2019-P-1.pdf (accessed on Sep 24, 2020).
5. Keating, B. A.; Carberry, P. S.; Bindraban, P. S.; Asseng, S.; Meinke, H.; Dixon, J. Eco-efficient Agriculture: Concepts, Challenges, and Opportunities. *Crop Sci.* **2010**, *50*, S-109, doi:10.2135/cropsci2009.10.0594.
6. Struik, P. C.; Kuyper, T. W. Sustainable intensification in agriculture: the richer shade of green. A review. *Agron. Sustain. Dev.* **2017**, *37*, 39, doi:10.1007/s13593-017-0445-7.
7. Tilman, D.; Balzer, C.; Hill, J.; Befort, B. L. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 20260–4, doi:10.1073/pnas.1116437108.
8. Srinivasan, A. *Handbook of Precision Agriculture: Principles and Applications*; 2006; ISBN 9781482277968.
9. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, doi:10.1016/j.rse.2019.111402.
10. Mulla, D. J. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* 2013, *114*, 358–371.
11. Patrício, D. I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81, doi:10.1016/j.compag.2018.08.001.
12. Gomes, J. F. S.; Leta, F. R. Applications of computer vision techniques in the agriculture and food industry: A review. *Eur. Food Res. Technol.* 2012, *235*, 989–1000.
13. Rahnemoonfar, M.; Sheppard, C. Deep count: Fruit counting based on deep simulated learning. *Sensors (Switzerland)* **2017**, *17*, 1–12, doi:10.3390/s17040905.
14. Aquino, A.; Millan, B.; Diago, M.-P.; Tardaguila, J. Automated early yield prediction in vineyards from on-the-go image acquisition. *Comput. Electron.*

Agric. **2018**, *144*, 26–36, doi:10.1016/j.compag.2017.11.026.

15. Aquino, A.; Diago, M. P.; Millán, B.; Tardáguila, J. A new methodology for estimating the grapevine-berry number per cluster using image analysis. *Biosyst. Eng.* **2017**, *156*, 80–95, doi:10.1016/j.biosystemseng.2016.12.011.

16. Pan, Z.; Huang, J.; Zhou, Q.; Wang, L.; Cheng, Y.; Zhang, H.; Blackburn, G. A.; Yan, J.; Liu, J. Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *34*, 188–197, doi:10.1016/j.jag.2014.08.011.

17. Cubero, S.; Aleixos, N.; Moltó, E.; Gómez-Sanchis, J.; Blasco, J. Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables. *Food Bioprocess Technol.* 2011, *4*, 487–504.

18. Zhang, B.; Huang, W.; Li, J.; Zhao, C.; Fan, S.; Wu, J.; Liu, C. Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Res. Int.* 2014, *62*, 326–343.

19. Moreda, G. P.; Ortiz-Cañavate, J.; García-Ramos, F. J.; Ruiz-Altisent, M. Non-destructive technologies for fruit and vegetable size determination - A review. *J. Food Eng.* **2009**, *92*, 119–136, doi:10.1016/j.jfoodeng.2008.11.004.

20. Sa'ad, F. S. A.; Ibrahim, M. F.; Shakaff, A. Y. M.; Zakaria, A.; Abdullah, M. Z. Shape and weight grading of mangoes using visible imaging. *Comput. Electron. Agric.* **2015**, *115*, 51–56, doi:10.1016/J.COMPAG.2015.05.006.

21. Sánchez Gómez, A. H.; García García, P.; Rejano Navarro, L. Elaboration of table olives. *Grasas y Aceites* **2006**, *57*, 86–94, doi:10.3989/gya.2006.v57.i1.24.

22. Kailis, S.; Harris, D. Producing Table Olives. **2007**, 342, doi:10.1017/S0014479707005662.

23. Aguilera, F.; Ruiz-Valenzuela, L. Forecasting olive crop yields based on long-term aerobiological data series and bioclimatic conditions for the southern Iberian Peninsula. *Spanish J. Agric. Res.* **2014**, *12*, 215–224, doi:10.5424/sjar/2014121-4532.

24. Bauer, M.; Cipra, J. Identification of Agricultural Crops by Computer Processing of ERTS MSS Data. *LARS Tech. Reports* **1973**.

25. R. O. Hoffman; D. M. Edwards; C. C. Eucker Identifying and Measuring Crop Type Using Satellite Imagery. *Trans. ASAE* **1976**, *19*, 1066–1070, doi:10.13031/2013.36177.

26. Zhang, C.; Kovacs, J. M. The application of small unmanned aerial systems for precision agriculture: A review. *Precis. Agric.* 2012, *13*, 693–712.

27. Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* 2014, *92*, 79–97.

28. Shih, F. Y. *Image Processing and Mathematical Morphology*; CRC Press, 2017; ISBN 9781138112285.

29. Soille, P. *Morphological Image Analysis : Principles and Applications*; Springer Berlin Heidelberg, 2004; ISBN 9783662050880.

30. Matsopoulos, G. K.; Marshall, S.; Brunt, J. N. H. Multiresolution morphological fusion of MR and CT images of the human brain. *IEE Proc. Vision, Image Signal Process.* **1994**, *141*, 137–142, doi:10.1049/ip-vis:19941184.

31. Zhao, Y. Q.; Gui, W. H.; Chen, Z. C.; Tang, J. T.; Li, L. Y. Medical images edge detection based on mathematical morphology. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*; 2005; Vol. 7 VOLS, pp. 6492–6495.

32. Coster, M.; Chermant, J. L. Image analysis and mathematical morphology for civil engineering materials. *Cem. Concr. Compos.* **2001**, *23*, 133–151, doi:10.1016/S0958-9465(00)00058-5.

33. Talbot, H.; Lee, T.; Jeulin, D.; Hanton, D.; Hobbs, L. W. Image analysis of insulation mineral fibres. *J. Microsc.* **2000**, *200*, 251–268, doi:10.1046/j.1365-2818.2000.00752.x.

34. Mueller, S.; Nickolay, B. Morphological image processing for the recognition of surface defects. In *Automated 3D and 2D Vision*; Ahlers, R.-J., Braggins, D. W., Kamerman, G. W., Eds.; SPIE, 1994; Vol. 2249, pp. 298–307.

35. Frydendal, I.; Jones, R. Segmentation of sugar beets using image and graph processing. In; Institute of Electrical and Electronics Engineers (IEEE), 2002; pp. 1697–1699.

36. Soille, P.; Pesaresi, M. Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2042–2055, doi:10.1109/TGRS.2002.804618.

37. Serra, J. *Image Analysis and Mathematical Morphology, vol. I*; Academic Press Inc.: Cambridge, MA, USA, 1982; ISBN 9780126372427.

38. Bojarski, M.; Del Testa, D.; Dworakowski, D.; Firner, B.; Flepp, B.; Goyal, P.; Jackel, L. D.; Monfort, M.; Muller, U.; Zhang, J.; Zhang, X.; Zhao, J.; Zieba, K. End to End Learning for Self-Driving Cars. **2016**.

39. Shin, H.-C.; Roth, H. R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R. M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298, doi:10.1109/TMI.2016.2528162.

40. Pereira, S.; Pinto, A.; Alves, V.; Silva, C. A. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1240–1251, doi:10.1109/TMI.2016.2538465.

41. Kamilaris, A.; Prenafeta-Boldú, F. X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90, doi:10.1016/j.compag.2018.02.016.

42. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **1980**, *36*, 193–202, doi:10.1007/BF00344251.

43. Hubel, D. H.; Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106–154, doi:10.1113/jphysiol.1962.sp006837.

44. Cun, L.; Cun, L.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; Jackel, L. D. Handwritten Digit Recognition with a Back-Propagation Network. *Adv. Neural Inf. Process. Syst.* **1990**, *2*, 396--404.

45. Raina, R.; Madhavan, A.; Ng, A. Y. Large-scale deep unsupervised learning using graphics processors. **2009**, 1–8, doi:10.1145/1553374.1553486.

46. Jia Deng; Wei Dong; Socher, R.; Li-Jia Li; Kai Li; Li Fei-Fei ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*; 2009; pp. 248–255.

47. Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. In *ImageNet Classification with Deep Convolutional Neural Networks*; 2012; pp. 1097–1105.

48. Szegedy, C.; Wei Liu; Yangqing Jia; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE, 2015; pp. 1–9.

49. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; 2016; Vol. 2016-Decem, pp. 2818–2826.

50. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; 2016; Vol. 2016-Decem, pp. 770–778.

51. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. **2016**.

52. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Int. Conf. Learn. Represent.* **2015**, 1–14.

List of Tables

Chapter 4. Results

List of Figures

Chapter 3. Materials

Chapter 4. Results