

**UNIVERSIDAD COMPLUTENSE DE MADRID**

FACULTAD DE CIENCIAS FÍSICAS

Departamento de Arquitectura de Computadores y Automática



**TESIS DOCTORAL**

**Métodos de visión por computador para localización e identificación de objetos y texturas en exteriores**

**Computer vision methods for objects and textures localization and identification in outdoors**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Juan Sebastián Romeo Granados**

Directores

Gonzalo Pajares Martinsanz  
Jesús Manuel de la Cruz García

**Madrid, 2014**

UNIVERSIDAD COMPLUTENSE DE MADRID  
Facultad de Ciencias Físicas



Departamento de Arquitectura de Computadores y Automática

## **TESIS DOCTORAL**

### ***Métodos de visión por computador para localización e identificación de objetos y texturas en exteriores***

Memoria para optar al grado de doctor presentada por:

D. Juan Sebastián Romeo Granados

Directores:

Dr. D. Gonzalo Pajares Martinsanz  
Dr. D. Jesús Manuel de la Cruz García

Madrid (España), 2014

# Agradecimientos

El trabajo de tesis que se presenta ha sido posible gracias a mi participación en los siguientes proyectos de investigación del Plan Nacional (por orden cronológico): *Sistema de Vigilancia, Búsqueda y Rescate en el Mar mediante colaboración de Vehículos Autónomos Marinos y Aéreos* (DPI2009-14552-C02-01); *PERcepción vIsual, toma de deCIsión y Actuación en sistemas artificiales complejos: control selectivo de malas hierbas en cultivos agrícolas* (PERICIA) (AGL2008-04670-C03-02/AGR) y en especial *Robot fleets for High Affective Agriculture and Forestry Management* (RHEA) financiado por la Comunidad Europea dentro del VII Programa Marco (FP7) con Grant Agreement NO.245986, dentro del THEME NMP-2009-3.4-1. Sin este soporte de investigación no habría sido posible la realización del trabajo que ahora se presenta. No resulta posible mencionar a todas las personas involucradas, por la larga lista de personas que la componen, ya que gran parte de ellas, en mayor o menor medida y directa o indirectamente han contribuido a este trabajo. Así pues vaya un agradecimiento general a todas aquellas personas que de alguna forma han contribuido a este trabajo.

Hay dos personas, no obstante, que no sólo han contribuido de forma directa a la realización de este trabajo sino que además me han sabido dar el apoyo académico, intelectual y personal necesario en los momentos oportunos. Me refiero a mis dos directores de tesis D. Gonzalo Pajares y D. Jesús M. de la Cruz, sin los cuales esta tarea habría resultado imposible.

Por último, no puedo dejar de agradecer, en el plano personal, a todas aquellas personas que me han apoyado con sus ánimos y admiración y por supuesto a mi familia por la paciencia que han demostrado, las horas extra que les ha tocado hacer y lo que me han echado de menos a lo largo de los años de trabajo para la elaboración de la Tesis en especial a Javier, Alejandro, Juan y Esther.

# ÍNDICE

## CAPÍTULO 1

INTRODUCCIÓN.....	1
1.1. Descripción del problema .....	1
1.2 Motivación.....	3
1.3 Objetivos.....	5
1.4 Metodología.....	6
1.5 Aportaciones de la investigación.....	7
1.5.1. Publicaciones en Revistas Indexadas. ....	8
1.5.2. Congresos y Conferencias. ....	10
1.5.3. Seminarios .....	12
1.5.4. Emisiones radiofónicas.....	12
1.6 Organización de la memoria de tesis.....	12

## CAPÍTULO 2

ALGORITMO PARA DETECCIÓN AUTOMÁTICA DE LÍNEAS DE CULTIVO ..	15
2.1 Introducción.....	15
2.2. Estado del arte .....	20
2.2.1. Métodos Basados en la Exploración de Líneas Horizontales.....	20
2.2.2. Métodos basados en la transformada de Hough .....	21
2.2.3. Métodos basados en el punto de fuga.....	23
2.2.4. Métodos basados en regresión lineal .....	23
2.2.5. Métodos basados en estereoscopia .....	23
2.2.6. Métodos basados en análisis de gotas (burbujas) .....	24
2.2.7. Métodos basados en la acumulación de plantas verdes.....	24
2.2.8. Métodos basados en análisis de frecuencia .....	25
2.3. Descripción del método .....	25
2.3.1. Imágenes Utilizadas.....	25
2.3.2. Extracción de Píxeles Verdes de la Imagen .....	28
2.3.3. Detección de Líneas .....	31
2.3.4. Restricciones.....	35
2.4. Resultados.....	37

## CAPÍTULO 3

CÁLCULO DE LAS <i>COORDENADAS X-Y</i> DE OBJETOS EN IMÁGENES .....	41
3.1. Introducción.....	41
3.2. Estado del arte .....	45
3.2.1. Estereopsis.....	46
3.2.2. Movimiento .....	47

3.2.3. Tonalidades.....	48
3.2.4. Texturas .....	48
3.2.5. Enfoque.....	49
3.2.6. Determinación de la profundidad con una sola cámara.....	49
3.3. Descripción del algoritmo .....	51
3.3.1. Cálculo de la <i>coordenada-y</i> .....	51
3.3.2. Cálculo de la <i>coordenada-x</i> .....	51
3.4. Resultados.....	53
3.4.1. <i>Coordenada-y</i> .....	53
3.4.2. <i>Coordenada-x</i> .....	55
 CAPÍTULO 4	
AJUSTES DE LA CÁMARA: PRECISIÓN EN LA DETECCIÓN DE LÍNEAS DE CULTIVO Y MALAS HIERBAS.....	61
4.1. Introducción.....	61
4.2. Parámetros extrínsecos .....	64
4.3. Factores de iluminación que afectan a la calidad de la imagen.....	67
4.4. Estado del arte .....	69
4.5. Materiales .....	70
4.6. Descripción del método .....	73
4.6.1. Parámetros extrínsecos .....	73
4.6.2. Factores de iluminación.....	75
4.7. Resultados.....	77
4.7.1. Análisis de los parámetros extrínsecos.....	78
4.7.2. Análisis de los factores de iluminación que afectan a la calidad de la imagen .....	83
 CAPÍTULO 5	
SISTEMA EXPERTO PARA LA IDENTIFICACIÓN DE VERDES EN IMÁGENES AGRÍCOLAS .....	87
5.1 Introducción.....	87
5.2. Estado del arte .....	88
5.3. Toma de decisiones .....	91
5.4. Identificación de verdes.....	93
5.4.1. Combinación de métodos clásicos.....	93
5.4.2. Preprocesado de imagen .....	95
5.4.3. Agrupamiento difuso <i>FC</i> .....	95
5.5. Resultados.....	96
5.5.1. Módulo de toma de decisiones: identificación de parámetros.....	97
5.5.2. Módulo de identificación de verde .....	102

5.5.3. Combinación de métodos clásicos.....	102
5.5.4. Preprocesamiento de imagen.....	103
5.5.5. Agrupamiento difuso <i>FC</i> .....	103
5.5.6. Método general: análisis cuantitativo.....	104
5.5.7. Método general: análisis cualitativo.....	106
5.5.8. Análisis de imágenes binarias.....	110
5.5.9. Análisis del agrupamiento difuso <i>FC</i> .....	111
CAPÍTULO 6	
CÁLCULO AUTOMÁTICO DEL UMBRAL DE PORCENTAJE DE VERDE .....	113
6.1. Introducción.....	113
6.2. Estado del arte.....	115
6.3. Descripción del método.....	115
6.4. Resultados.....	117
CAPÍTULO 7	
CONCLUSIONES.....	123
7.1. Algoritmo para la detección automática de líneas de cultivo.....	124
7.2. Cálculo de las <i>coordenadas x-y</i> de objetos en imágenes.....	124
7.3. Ajustes de la cámara para detección de líneas de cultivo y malas hierbas con la máxima precisión.....	125
7.4. Sistema experto para identificación de verdes en imágenes agrícolas.....	126
7.5. Algoritmo para el cálculo automático del umbral de porcentaje de verde.....	127
CAPÍTULO 8	
TRABAJOS FUTUROS.....	129
8.1. Mejora de los métodos y algoritmos propuestos.....	129
8.2. Adaptación de los métodos para su utilización en otras tareas.....	130
REFERENCIAS.....	233



## INTRODUCCIÓN

### 1.1. Descripción del problema

Los avances tecnológicos en el ámbito de la visión computacional permiten, y sin duda permitirán en el futuro, la aplicación de técnicas basadas en visión artificial a diferentes áreas tanto de producción como de investigación. Desde sus inicios, la visión por computador ha resultado ser una herramienta útil en diversos campos aunque sólo sea por hacer bueno el dicho popular de que “más vale una imagen que mil palabras”.

La comunidad científica distingue al menos dos tipos de entornos claramente diferenciados en los que se pueden aplicar técnicas de visión por computador, a saber: *a) Entornos estructurados* y *b) Entornos no estructurados*.

Los entornos estructurados, principalmente ubicados en el ámbito de producción industrial, presentan la peculiaridad de que muchas de las variables externas involucradas en los procesos están perfectamente controladas. A este tipo de entornos pertenecen, por ejemplo, los sistemas de control de calidad en un proceso de fabricación industrial de determinado tipo de elementos o piezas en una fábrica. Para entender el concepto de entorno estructurado, pensemos en una cadena de producción de piezas, donde éstas se presentan sobre un determinado panel o cinta transportadora en la que el fondo y las piezas se distinguen con claridad, la velocidad de la cinta es conocida, generalmente constante, y la iluminación está producida por un sistema fijo con una dirección, intensidad y naturaleza controlada, además de conocida. Los métodos de procesamiento de imágenes solamente deben atender a las variables intrínsecas de la aplicación, mientras que las extrínsecas se encuentran perfectamente controladas.

Por el contrario, en los entornos no estructurados las variables externas antes mencionadas, u otras que aparecen de forma espontánea, no están sometidas a ningún

tipo de control, presentando, por lo general, una aleatoriedad de especial relevancia. Los entornos naturales pertenecen a esta categoría. Por entornos naturales nos referimos a aquellos en los que el sistema de visión se instala y procesa las imágenes procedentes de exteriores. Las aplicaciones basadas en sistemas de visión para agricultura o en entornos acuáticos, bien pantanos o mar abierto, son dos ejemplos claros de entornos no estructurados. Las condiciones ambientales, altamente cambiantes, hacen que los métodos de procesamiento y tratamiento de imágenes deban diseñarse con suficiente robustez para su adaptación a dichas condiciones y por lo general abordando el problema desde puntos de vista completamente distintos a como se haría en entornos controlados. A modo de ejemplo, supongamos un tractor robotizado y equipado con un sistema de visión para la realización de tareas propias en Agricultura de Precisión, pongamos por caso detección de líneas de cultivo para guiado e identificación de malas hierbas para su tratamiento selectivo. El tractor realiza estas tareas durante diferentes días y en diferentes momentos durante un mismo día. Resulta obvio pensar que las condiciones ambientales varían de un momento a otro, principalmente debido a la existencia de distintas condiciones ambientales, tal es el caso de días nublados y soleados, del tránsito de nubes que producen sombras y áreas más iluminadas en espacios muy breves de tiempo o de la incidencia de la luz solar desde diferentes ángulos: frontal, lateral, posterior. Además, en este tipo de ambientes, existe una alta y cambiante radiación infrarroja y ultravioleta afectando directamente a la calidad de las imágenes. Estas circunstancias también ocurren en los entornos que involucran superficies de agua (pantanos o áreas marinas) donde además de lo anterior, aparecen reflejos en la superficie del agua que interfieren desfavorablemente en los procesos de visión.

Resulta evidente que los procesos de imágenes en los entornos no estructurados requieren el diseño de métodos robustos y eficaces para hacer frente a los inconvenientes reseñados anteriormente. Es justamente en esta línea donde se centra la investigación que se presenta en esta memoria de Tesis. Concretamente, se proponen una serie de métodos y procedimientos encaminados al tratamiento de imágenes en entornos no estructurados, con fines de aplicación específicos, cuya concreción se formula a lo largo de los diferentes capítulos que dan cuerpo a esta Tesis.

## 1.2 Motivación

El trabajo de investigación desarrollado tiene su origen en los proyectos de investigación que se mencionan a continuación, todos ellos dentro del marco de investigación en el grupo ISCAR (2013), oficialmente reconocido por la Universidad Complutense de Madrid desde el año 2006. Algunos de ellos en colaboración con otros grupos de investigación tanto de carácter nacional como internacional, tal y como se refleja en las descripciones siguientes:

1. La actividad investigadora se inicia en el marco del proyecto del Plan Nacional de I+D 2008 para el trienio 2009/2011 concedido con el número de referencia AGL2008-04670-C03-02/AGR (AGL, 2008) con el título *PERcepción vIsual, toma de deCIsión y Actuación en sistemas artificiales complejos: control selectivo de malas hierbas en cultivos agrícolas (PERICIA)*. En este proyecto, mediante un contrato de trabajo asociado al mismo, se investigaron técnicas de visión estereoscópica para la localización espacial de rodales de malas hierbas en campos de cultivo de cereal y maíz con el fin de proceder a su pulverización selectiva. El objetivo final del proyecto consistía en minimizar el efecto contaminante medioambiental por el tratamiento con herbicidas de los campos de cultivo, así como disminuir los costes de producción. Este proyecto ha tenido su continuidad en el nuevo proyecto titulado *Sistema de inspección terrestre en vehículos autónomos y su aplicación efectiva a la detección de malas hierbas y su control localizado* (AGL2011-30442-C02-02), perteneciente igualmente al Plan Nacional de I+D+i para el trienio 2011/2014, AGL (2011). Se trata de un subproyecto coordinado en colaboración con el Centro de Automática y Robótica, centro mixto CSIC-Universidad Politécnica de Madrid, así como con el Instituto de Agricultura Sostenible, también del CSIC. El objetivo principal estriba en la localización de las malas hierbas dentro de los campos de cultivo de maíz y cereal. Como en el caso anterior, la finalidad es idéntica a la planteada en su predecesor, si bien las técnicas relativas al procesamiento de imágenes se centran más en la localización espacial de estructuras que en la identificación.
2. Proyecto del Plan Nacional de I+D+i titulado *Sistema de Vigilancia, Búsqueda y Rescate en el Mar mediante colaboración de Vehículos Autónomos Marinos y Aéreos* (DPI2009-14552-C02-01), *DPI (2009)*, prorrogado durante un año más

al previsto en su duración inicial. En este proyecto se plantea como una de las tareas fundamentales la colaboración de vehículos tanto aéreos como terrestres y marinos. En todos los casos se hace necesario establecer el posicionamiento relativo entre ellos, siendo de especial relevancia el posicionamiento de barcos para determinar sus coordenadas geográficas a partir de un punto de referencia que establece las coordenadas de base o referencia. Este tipo de localización se lleva a cabo con técnicas de visión artificial mediante el calibrado de la cámara para posiciones fijas y relevantes en el entorno de trabajo. Los experimentos se desarrollan en un entorno acuático, concretamente en el embalse de Valmayor (Madrid), gestionado por el Canal de Isabel II.

3. Proyecto financiado por la Comunidad Europea dentro del VII Programa Marco (FP7) cuya duración abarca el cuatrienio 2010-2014, con el título *Robot fleets for High Affective Agriculture and Forestry Management (RHEA)*, incluido en el THEME NMP-2009-3.4-1 (*RHEA 2014*). Este proyecto se desarrolla bajo un consorcio constituido por diecinueve centros de investigación y empresas pertenecientes a ocho países europeos. El objetivo consiste en diseñar tres prototipos constituidos en una flota de tractores robotizados y coordinados, conjuntamente con vehículos aéreos autónomos del tipo conocido como “drones”. Cada uno de los prototipos robotizados tiene prevista su actuación en tres campos de cultivo, a saber: maíz, cereal y olivar. Los tres tractores se encuentran sensorizados con el objetivo de realizar navegación autónoma y localización de las áreas de interés, también de forma autónoma, para tratamientos selectivos. Es en el tractor dedicado al maíz en el que se centra la actividad del presente trabajo de investigación. Este tractor dispone de un equipo de visión instalado en su parte superior mediante el cual se llevan a cabo procesos tanto de identificación de líneas de cultivo como de densidades de malas hierbas entre las citadas líneas (*figura 1.1*). Aquí es donde se centra el objetivo principal de la investigación llevada a cabo en el presente trabajo. La investigación desarrollada en este proyecto se desarrolla mediante un contrato de trabajo asociado al mismo.



*Figura 1.1: Tractor robotizado equipado con sistema de visión en su parte superior.*

Queda patente, a través de la breve descripción de los proyectos mencionados, la existencia de entornos no estructurados, originarios de las técnicas de visión por computador que se describen en el presente trabajo. El desarrollo de métodos eficientes de tratamiento de imágenes para tales entornos y sistemas, constituye la principal motivación del trabajo de investigación realizado.

Desde el punto de vista tecnológico, la motivación viene suscitada por el hecho de automatizar una serie de tareas orientadas a conseguir con la mayor eficacia y eficiencia los objetivos establecidos para los vehículos autónomos dotados con los sistemas de visión apropiados.

### **1.3 Objetivos**

A la vista de las consideraciones expresadas en los apartados precedentes, se plantean los siguientes objetivos de investigación relativos a los procesos de segmentación y localización previstos:

1. Diseñar estrategias de segmentación de imágenes en entornos *no estructurados*, con el fin de identificar estructuras relevantes en dichos entornos.

2. Diseñar los métodos de identificación de elementos relevantes en las imágenes. En las aplicaciones agrícolas son líneas de cultivo y rodales de malas hierbas con fines de tratamientos selectivos. En la aplicación involucrando superficies de agua, son barcos autónomos navegando sobre dicha superficie.
3. Identificar las líneas de investigación futuras, tanto desde el punto de vista de la mejora de las estrategias propuestas como desde el diseño de nuevos métodos para su aplicación a otros entornos no estructurados.
4. Divulgar y difundir los resultados de la investigación mediante publicaciones científicas u otros mecanismos.

Conviene destacar el hecho de que todos los métodos propuestos han sido ideados y desarrollados para su funcionamiento en tiempo real. Esto conlleva una doble dificultad debido a dos razones muy evidentes, a saber:

- a. Trabajar en tiempo real limita de forma considerable la capacidad de procesamiento debido al ajustado tiempo de cómputo disponible. Esto implica el desarrollo de algoritmos realmente optimizados sin que por ello pierdan robustez y eficacia.
- b. Los entornos no estructurados obligan a constantes ajustes de parámetros y variables propias de los algoritmos utilizados. Al tratarse de algoritmos diseñados para tiempo real, estos ajustes de variables y sus calibraciones o mediciones previas, necesarias para determinar los valores apropiados, también han de estar automatizados. Esto conlleva no sólo el desarrollo de los algoritmos específicos para los objetivos marcados sino también el desarrollo de algoritmos particulares (algoritmos secundarios), necesarios para ajustar permanentemente de forma automática los parámetros requeridos por el algoritmo principal.

## **1.4 Metodología**

En el desarrollo del trabajo se han seguido las directrices marcadas por el método científico. Inicialmente, se ha realizado una búsqueda de información bibliográfica en todas las fuentes disponibles, utilizando principalmente para ello el

acceso a través de la red del Consorcio Madroño al que están suscritas las Universidades de la Comunidad de Madrid. Se han estudiado los antecedentes metodológicos y se ha analizado y evaluado el estado del arte en este área. Seguidamente, se han dado los pasos que se indican a continuación, a la vez que se han revisado de manera continua las fuentes bibliográficas hasta el final de la investigación e incluso en el momento de la redacción de este trabajo. Las etapas seguidas han sido:

1. Definición del problema de investigación en función tanto de las necesidades identificadas como de la motivación establecida previamente.
2. Planteamiento de los objetivos de la investigación.
3. Formulación de planteamientos verificables.
4. Obtención de datos experimentales.
5. Clasificación, análisis e interpretación de los datos.
6. Conclusiones generales.
7. Descripción de líneas futuras de investigación.

Para la realización de los experimentos se han implementado los métodos basados en las técnicas propuestas. Este proceso abarca desde la adquisición de datos para la creación de una base suficientemente amplia con la que experimentar, hasta la obtención de resultados, pasando por su visualización.

## **1.5 Aportaciones de la investigación**

La contribución principal de este trabajo consiste en el desarrollo de una serie de técnicas de segmentación de imágenes en entornos no estructurados y la posterior identificación de estructuras relevantes. En concreto, se describen cuatro de los algoritmos utilizados en los proyectos mencionados y un estudio realizado, con resultados concretos, para la mejora de la precisión de los resultados. La materialización de los trabajos de investigación se resume como sigue, algunos de los cuales han permitido su divulgación internacional, tal y como se especifica en las siguientes secciones.

### 1.5.1. Publicaciones en Revistas Indexadas.

1. **Juan Romeo**, G. Pajares, M. Montalvo, J. M. Guerrero, M. Guijarro, and A. Ribeiro (2012). “*Crop Row Detection in Maize Fields Inspired on the Human Visual Perception*”. The ScientificWorld Journal, vol. 2012, Article ID 484390, 10 pages.

Se plantea una propuesta para la identificación de las líneas de cultivo en campos de maíz, inspirada en el proceso de reconocimiento humano. Es un algoritmo novedoso ideado para trabajar en tiempo real y en entornos no estructurados. Destaca por su robustez en la detección de líneas, robustez que se materializa y verifica frente a movimientos del tractor y a la presencia masiva de malas hierbas e incluso frente a la ausencia parcial de líneas de cultivo. Las líneas se identifican buscando acumulaciones máximas de píxeles verdes donde supuestamente se ubican las líneas de cultivo buscadas con márgenes suficientemente amplios. Para la extracción de aquellos píxeles que pertenecen a plantas, considerados con predominio de verde, se aplica una técnica basada en lógica borrosa. Este método completo se describe en el *capítulo 2: Detección de líneas de cultivo*.

2. **Juan Romeo**, J. M. Guerrero, M. Montalvo, L. Emmi, M. Guijarro, P. Gonzalez-de-Santos, G. Pajares (2013). “*Camera Sensor Arrangement for Crop/Weed Detection Accuracy in Agronomic Images*”. Sensors, 13, 4348-4366.

Se identifican los parámetros extrínsecos más relevantes para una mayor resolución de las imágenes obtenidas desde la cámara ubicada en el tractor. Una vez identificados, se calculan empíricamente los valores óptimos.

Finalmente y por otro lado, se describen dos métodos de rectificación de las imágenes. Uno para compensar los efectos de la iluminación cambiante controlando por programa el tiempo de exposición en la toma de imagen y otro para corregir el efecto conocido como vignetting consistente en la aparición de un anillo difuso centrado en la imagen provocado por el uso de filtros ópticos. Este método se describe en el capítulo cuatro: Ajustes de la cámara: precisión en la detección de líneas de cultivo y malas hierbas.

3. **Juan Romeo**, G. Pajares, M. Montalvo, J.M. Guerrero, M. Guijarro, J.M. de la Cruz. (2013). “*A new Expert System for greenness identification in agricultural images*”. *Expert Systems with Application*, 40, 2275–2286.

Se propone un sistema experto para la segmentación de imágenes y la obtención de píxeles verdes existentes en las mismas. Para ello, se combinan distintos métodos de segmentación y análisis de imágenes de una forma coordinada y complementaria. El resultado es la aplicación de un criterio lógico de actuación en el procesamiento de la imagen en función de la calidad inicial de éstas, que conduce a la identificación de las partes correspondientes a la vegetación para su diferenciación con el resto de las partes. El método se describe en el capítulo 5: Sistema experto para la identificación de verdes en imágenes agrícolas.

4. J. M. Guerrero, G. Pajares, M. Montalvo, **Juan Romeo**, M. Guijarro (2012). “*Support Vector Machines for crop/weeds identification in maize fields*”. *Expert Systems with Applications*, 39(12), 11149-11155.

Trabajo centrado en el estudio y análisis comparativo de uno de los métodos de binarización utilizado para comprobar la eficacia del procedimiento propuesto en el artículo en relación a la binarización de la imagen.

5. M. Montalvo, G. Pajares, J. M. Guerrero, **Juan Romeo**, M. Guijarro, Á. Ribeiro, J.J. Ruz, J. M. de la Cruz (2012). “*Automatic detection of crop rows in maize fields with high weeds pressure*”. *Expert Systems with Applications*, 39(15), 11889-11897.

Trabajo centrado en el análisis y comparación con otras técnicas del método propuesto en el artículo en relación a la técnica descrita en la referencia uno, reseñada previamente.

6. J.M. Guerrero, M. Guijarro, M. Montalvo, **Juan Romeo**, L. Emmi, Á. Ribeiro, G. Pajares (2013). “*Automatic expert system based on images for accuracy crop row detection in maize fields*”. *Expert Systems with Applications*, 40(2), 656–664.

Trabajo centrado en el análisis y estudio sobre la precisión del método propuesto en el artículo en relación a la técnica descrita en la referencia uno.

7. M. Montalvo, J.M. Guerrero, **Juan Romeo**, L. Emmi, M. Guijarro, G. Pajares (2013). “*Automatic expert system for weeds/crops identification in image from maize fields*”. *Expert Systems with Applications*, 40(1), 75–82.

Trabajo basado en la ayuda a la identificación de verdes procedentes de plantas de maíz y malas hierbas llevada a cabo en el trabajo de la referencia tres.

### 1.5.2. Congresos y Conferencias.

8. **Juan Romeo**, D. Oliva, J. M. Guerrero, M. Montalvo, M. Guijarro, G. Pajares (2012). “*Plants discrimination in maize fields applying human perception to digital images*”. *Proc. Int. Conf. of Agricultural Engineering (CIGR-AgEng2012)*, July 8-12, Valencia (Spain), (<http://cigr.ageng2012.org>).

Se propone un nuevo método para “cuantificar la diferencia cualitativa entre dos colores”. Es una complejión de la distancia de Manhattan que corrige los errores de ésta cuando la diferencia entre dos colores no es una diferencia equilibrada en los tres canales espectrales.

9. J. M. Guerrero, **Juan Romeo**, L. Emmi, M. Montalvo, M. Guijarro, G. Pajares, P. González-de-Santos (2012). “*Influence of the vision system pitch angle on crop and weeds detection accuracy*”. *First RHEA Int. Conf. on Robotics and associated High-technologies and Equipment for Agriculture*, September 19-21, Pisa (Italy), (<http://www.rhea-conference.eu/2012/>).

Precursor de una parte del trabajo descrito en la referencia dos. Se expone un estudio sobre la influencia del ángulo de inclinación en la resolución de las imágenes y la precisión de cálculo con éstas. Parte de este método se encuentra descrito en el capítulo cuatro.

10. J. M. Guerrero, M. Montalvo, **Juan Romeo**, D. Oliva, M. Guijarro, G. Pajares (2012). “*Detection of images with high CCD charge level due to indirect sunlight in maize fields*”. *Int. Conf. of Agricultural Engineering (CIGR-AgEng2012)*, July 8-12, Valencia (Spain), (<http://cigr.ageng2012.org>).

Se estudian métodos de compensación de la entrada de luz al sensor, en parte también recogido en el trabajo descrito en la referencia dos.

11. M. Montalvo, J. M. Guerrero, **Juan Romeo**, D. Oliva, M. Guijarro, G. Pajares (2012). “*Unsupervised learning for crop/weeds discrimination in maize fields with high weed densities*”. Int. Conf. of Agricultural Engineering (CIGR-AgEng2012), 5 pages, July 8-12, Valencia (Spain), (<http://cigr.ageng2012.org>).

Se estudian y analizan métodos de identificación de verdes para su diferenciación del resto de la imagen. Trabajo desarrollado conjuntamente con el descrito en la referencia tres.

12. M. Montalvo, J.M. Guerrero, **Juan Romeo**, M. Guijarro, J.M. de la Cruz, G. Pajares (2013). “*Acquisition of Agronomic Images with Sufficient Quality by Automatic Exposure Time Control and Histogram Matching*”. Advanced Concepts for Intelligent Vision Systems (ACIVS'13) (J. Blanc-Talon, A. Kasinski, W. Philips, D. Popescu, P. Scheunders, Eds.), Lecture Notes in Computer Science, vol. 8192, pp. 37-48, Springer International Publishing Switzerland. October 28-31, Poznan, Poland.

Corresponde al estudio y análisis de la calidad de las imágenes mediante la variación de los tiempos de exposición del sensor CCD. En parte es el resultado de las técnicas propuestas en la referencia dos.

13. M. Guijarro, J.M. Guerrero, M. Montalvo, **Juan Romeo**, G. Pajares (2013). “*A new approach to solve image thresholding in precision agriculture*”. 9th European Conference on Precision Agriculture. Lérida, Spain, July 7-11, poster. <http://www.ecpa2013.udl.cat/>

Estudio comparativo de métodos de binarización para identificación de verdes en las imágenes.

14. M. Montalvo, J.M. Guerrero, **Juan Romeo**, P.J. Herrera, M. Guijarro, A. Ribeiro (2011). “*Camera System geometry for site specific treatment in precision agriculture*”. in Proc. First RHEA Workshop (G. Rabatel and P. González de Santos), September 9, Montpellier (France).

Estudio y definición de la geometría del sistema de visión. Precursor del trabajo descrito en la referencia dos, donde se exponen diversos aspectos relacionados con la disposición de las cámaras y los sistemas de referencia asociados.

### 1.5.3. Seminarios

15. Conferencia en la Facultad de informática de la Universidad Complutense de Madrid (mayo 2011), titulada “*Sistemas de visión estereoscópica para control de robots de exterior y en agricultura de precisión*”, en el marco de convenio de colaboración del Fondo de Cooperación Internacional en Ciencia y Tecnología Unión Europea-México (FONCICYT). Con la participación del Instituto Politécnico Nacional, la Universidad de Guadalajara (México), Universidad Libre de Berlín y Universidad Complutense de Madrid.

16. Conferencia titulada “*Aplicaciones prácticas de la visión artificial en exteriores*”, de la asignatura “Seminarios de visión artificial” dentro del Master de Visión Artificial de la Universidad Rey Juan Carlos I, Febrero 2012 (2 horas).

### 1.5.4. Emisiones radiofónicas.

17. Participación en el programa de radio SIN DISTANCIAS- titulado “Ojos artificiales en agricultura” emitido el 21 de marzo 2012 dentro de la programación de la UNED en RADIO-3 de RNE.

## 1.6 Organización de la memoria de tesis

La memoria de tesis se presenta estructurada en ocho capítulos, que se organizan según la siguiente distribución:

**Primer capítulo.** *Introducción*, donde se describe el problema y se estudian y plantean los objetivos de la tesis, así como la motivación e introducción de las propuestas y planteamientos realizados, tal y como se ha expuesto en las secciones precedentes.

**Segundo capítulo.** *Algoritmo para la detección automática de líneas de cultivo*, donde se explica en detalle los pasos realizados para la detección de las líneas de cultivo en plantaciones de maíz, a partir de las imágenes obtenidas por la cámara ubicada en el tractor. Como ya se ha mencionado anteriormente, dicho algoritmo fue publicado en *Romeo y col. (2012)*

**Tercer capítulo.** *Cálculo de las coordenadas x-y de objetos en imágenes*, donde se explica un método para ubicar espacialmente objetos a partir de una única imagen.

**Cuarto capítulo.** *Ajustes de la cámara: precisión en la detección de líneas de cultivo y malas hierbas*, el cual presenta un estudio realizado para determinar la posición óptima de la cámara (altura, ángulo de inclinación y distancia de enfoque) con el objeto de obtener la máxima precisión en los algoritmos posteriores de procesamiento de imágenes. Igualmente, este método ha sido publicado en *Romeo y col. (2013a)*.

**Quinto capítulo.** *Sistema experto para la identificación de verdes en imágenes agrícolas*, presenta un método de identificación de verdes en imágenes agrícolas basado en distintas etapas donde según la calidad de las imágenes, se aplican unos u otros algoritmos de procesamiento para la obtención de los mencionados píxeles verdes con la máxima precisión posible. Nuevamente, el presente método ha sido publicado en *Romeo y col. (2013b)*.

**Sexto capítulo.** *Algoritmo para el cálculo automático del umbral de porcentaje de verde* es un sencillo e ingenioso método de autoajustar el porcentaje de verdes de forma automática y en tiempo real a las imágenes para su segmentación.

**Séptimo capítulo.** *Conclusiones*, donde se exponen las conclusiones más relevantes de los trabajos realizados y explicados detalladamente en los capítulos anteriores.

**Octavo capítulo.** *Trabajos futuros*, dado que todo trabajo es siempre susceptible de mejoras y optimizaciones posteriores, en este capítulo se proporcionan algunas líneas de trabajo posibles para ello.



# ALGORITMO PARA DETECCIÓN AUTOMÁTICA DE LÍNEAS DE CULTIVO

## 2.1 Introducción

El presente método ha sido ideado y desarrollado dentro del proyecto RHEA (*RHEA 2014*) perteneciente al VII Programa Marco de la Unión Europea en nanociencias y nanotecnología. Algunos estudios preliminares se desarrollaron en el proyecto AGL (2008). Los resultados iniciales se publicaron en *Romeo y col. (2012a)*.

En dicho proyecto un tractor autónomo (no tripulado) se desplaza a lo largo de las líneas de cultivo de maíz detectando malas hierbas para realizar un tratamiento selectivo.

La detección de las malas hierbas se lleva a cabo mediante técnicas de visión artificial gracias a una cámara colocada en la parte frontal del tractor que va capturando y procesando imágenes en tiempo real. Una vez procesadas dichas imágenes, las malas hierbas quedan detectadas, localizadas y cuantificadas en términos de porcentaje de verde.

La distinción entre plantas de cultivo y malas hierbas es una tarea ciertamente complicada debido a que ambas presentan componentes espectrales muy similares. Así pues, no se pueden separar plantas de malas hierbas con métodos basados exclusivamente en dichas componentes espectrales, si bien es cierto que algunos filtros de banda estrecha para el NIR permiten en otros casos, distinguir algunas malas hierbas de cultivos. Dicha distinción se realiza detectando las líneas de cultivo y considerando como mala hierba cualquier vegetación que esté fuera de las mismas. Se asume que la presencia de malas hierbas entre líneas de cultivo implica igualmente la presencia de

malas hierbas entre las plantas de maíz en esa misma ubicación procediéndose por tanto a su tratamiento específico.

Debido a que el tractor navega en un terreno de cultivo, lleno de irregularidades provocadas por piedras, baches, surcos, etc., se producen vibraciones y pequeñas variaciones de los ángulos de inclinación, orientación y alabeo de la cámara. Esto hace que en cada imagen, las líneas de cultivo aparezcan en posiciones ligeramente desplazadas y/o giradas respecto de la imagen anterior.

Por ello, la detección de las líneas de cultivo es una tarea crítica en el proyecto RHEA, tanto en precisión como en velocidad de cálculo ya que ha de llevarse a cabo en tiempo real para ajustarse a la velocidad de navegación del tractor, que varía entre 3 y 6 Km/h.

Existen dos razones fundamentales por las que es necesario detectar las líneas de cultivo:

- Por un lado, como acabamos de mencionar, es un paso previo necesario para delimitar y ubicar las zonas de malas hierbas dentro de la imagen. Debido a que una vez detectadas las líneas de cultivo, se establecen unos márgenes alrededor de las mismas para abarcar el ancho de las plantas de maíz. Estos márgenes serán mayores cuanto mayor sea el estado de crecimiento en el momento del tratamiento. Cualquier vegetación que quede fuera de estos márgenes se considera mala hierba. Para la localización espacial de estas malas hierbas se establecen sobre la imagen una serie de cuadrículas o celdas que delimitan la zona a tratar.
- Por otro, es igualmente necesario para el correcto guiado del tractor durante su marcha siguiendo las líneas de cultivo. Cualquier desviación del tractor respecto a su trayectoria programada y considerada válida puede ser detectado mediante la variación de las líneas de cultivo en la imagen respecto de la posición teóricamente correcta que éstas deberían tener cuando el tractor se encuentra correctamente alineado. De esta forma podemos proporcionar las distancias y ángulos de desviación necesarios para su correcto guiado.

Las imágenes utilizadas han sido tomadas desde la cámara colocada en el arco frontal del tractor específicamente dedicado al cultivo de maíz dentro del proyecto RHEA (RHEA 2014), *figura 2.1*. La posición de la cámara es tal que apunta en la dirección de la marcha del tractor con un cierto ángulo de inclinación respecto del suelo y enfocando siempre hacia la parte frontal del tractor como referencia. En esta parte frontal se ha colocado un panel con los tres colores primarios RGB que nos permite calibrar la calidad de los colores en cada imagen obtenida, a la vez que comprobar que la cámara está captando imágenes correctamente por el hecho de que en las posiciones esperadas aparecen los colores previstos. Por otro lado, dicho panel constituye una referencia fija de la posición del tractor respecto de las líneas de cultivo. En caso de que el tractor modificase su trayectoria, su desviación podría ser rápidamente identificada y cuantificada gracias a las referencias estáticas con respecto al panel frontal. Igualmente, cualquier desajuste de la posición de la cámara en el tractor, debido a golpes inesperados, vibraciones u oscilaciones, es fácilmente controlado al tener la referencia fija de dicho panel en todas las imágenes, *figuras 2.2(a)* y *2.2 (b)*. Obsérvese en estas imágenes cómo la referencia del frontal del tractor es necesaria e imprescindible para poder diferenciar ambas circunstancias dado que la disposición de las líneas en ambos casos es idéntica. En ambas imágenes se observa cómo una desviación del tractor con respecto a las líneas de cultivo se manifiesta en forma de desajuste del panel con respecto a dichas líneas, lo que permite establecer primero que se ha producido una desviación y segundo la magnitud de dicha desviación.



**Figura 2.1:** Ubicación de la cámara en la parte frontal del tractor centrada sobre el arco elevado.



**Figura 2.2:** (a) Imagen obtenida por la cámara al desviarse el tractor de su trayectoria.  
(b) Imagen obtenida por la cámara al moverse la cámara respecto del tractor.

Las imágenes obtenidas con la cámara del tractor tienen el aspecto mostrado en la *figura 2.3*, donde la profundidad alcanzada en la imagen depende del ángulo de inclinación con el que se coloque la cámara, así como de los movimientos del tractor provocados por la orografía del terreno. Cada una de las imágenes mostradas en dicha figura está tomada con un ángulo de inclinación de la cámara diferente con el fin de verificar distintas posiciones a modo de prueba.



**Figura 2.3:** Distintas imágenes de líneas de cultivos tomadas desde el tractor con el eje óptico de la cámara paralelo a la dirección de la marcha y con distintos ángulos de inclinación.

El método descrito en esta sección es válido para cualquier imagen con líneas de cultivo siempre y cuando estén tomadas con la proyección mostrada, es decir, con proyección de perspectiva, independientemente de la altura a la que se encuentre la cámara o la inclinación de la misma con respecto al suelo.

La única condición para poder aplicar el presente algoritmo es que las líneas de cultivo que se quieren detectar deben “nacer” en la parte inferior de la imagen y converger según nos vamos aproximando a la parte superior de la imagen. No es necesario que las líneas lleguen a unirse en el punto de fuga (punto de convergencia en

perspectiva) al llegar a la parte superior de la imagen, es decir, basta con que se obtengan con proyección de perspectiva.

El presente algoritmo está ideado con la intención de discriminar las malas hierbas de las líneas de cultivo en plantaciones de maíz para su posterior tratamiento selectivo. Esto significa que las líneas de cultivo han de ser identificadas y localizadas con la mayor precisión posible. Algunos de los requisitos propuestos por *Astrand y Baerveldt (2005)* y explicados por *Slaughter y col. (2008)* para sistemas de guiado son, en esencia, muy parecidos y pueden tenerse en cuenta en la detección de líneas de cultivo en los métodos diseñados a tal fin como el que nos ocupa. Así pues nuestro sistema está diseñado con los siguientes objetivos:

- Localizar las líneas de cultivo con la mayor precisión posible.
- Trabajar en tiempo real.
- Debe ser capaz de procesar las imágenes en campos de cultivo con malas hierbas y en estados tempranos de crecimiento donde las alturas y diferencias espectrales entre malas hierbas y cultivos son prácticamente inapreciables.
- Fiable con líneas de cultivo que presenten zonas sin germinar (zonas en blanco a lo largo de las líneas).
- Independiente a una presencia elevada de malas hierbas.
- Robusto ante las inclemencias del tiempo atmosférico (variación de luminosidad, radiación infrarroja, sombras, reflejos, etc.).
- Debe diseñarse con las mínimas restricciones posibles para abarcar el mayor número de situaciones posibles.

En resumen, el objetivo pretendido con el presente algoritmo es desarrollar un método general y práctico para la identificación de las líneas de cultivo en campos de maíz, en tiempo real, a partir de imágenes captadas por una cámara instalada a bordo del tractor, que navega de forma autónoma en el campo de cultivo. Para ello, hacemos uso de las ventajas proporcionadas por los diferentes métodos existentes que describimos a

continuación extrayendo las ideas expuestas y diseñando una nueva estrategia para la detección inspirada en la percepción de la visión humana.

## 2.2. Estado del arte

La detección de las líneas de cultivo es un problema que aparece continuamente en tareas agrícolas automatizadas y realizadas con tractores agrícolas robotizados y equipados con sistemas de visión. En Agricultura de Precisión, estas tareas cobran especial relevancia. El tractor navega y actúa sobre un área específica dentro de superficies más extensas (*Davies y col. 2013*), el sistema de percepción visual proporciona una parte importante de la información requerida tanto para navegación como para detección de elementos de interés en la imagen (*Rovira-Mas y col., 2010*).

Resulta bien conocido que un importante asunto relacionado con la aplicación de métodos basados en Visión por Computador se refiere a la detección tanto de líneas de cultivo como de malas hierbas presentes en los campos de cultivo, por ejemplo maíz. Estos temas han sido ampliamente tratados en la literatura (*Onyango y Marchant, 2003; Burgos-Artizzu y col., 2009; Tellaeché y col., 2008a-b; Sainz-Costa y col., 2011*). Gracias a las técnicas desarrolladas en los diferentes trabajos es posible la aplicación de tratamientos específicos con el fin de eliminar las malas hierbas en favor del crecimiento de los cultivos.

Veamos a continuación una breve descripción de los métodos propuestos para la detección de líneas de cultivo basados en distintas estrategias.

### 2.2.1. Métodos Basados en la Exploración de Líneas Horizontales

*Søgaard y Olsen (2003)*, aplicaron transformación del color RGB a la escala de grises. Esto se hizo dividiendo primero la imagen en sus tres canales espectrales (rojo, verde y azul) y después aplicando métodos conocidos y fiables para extraer las partes pertenecientes a plantas en *Woebbbcke (1995)*. Después de esto, la imagen de grises se divide en tiras horizontales donde los valores máximos de gris indican la presencia de una fila candidata. Cada máximo determina un segmento de fila y el centro de gravedad del segmento aparece marcado en la posición de la tira. Las líneas de cultivo son identificadas uniendo los puntos marcados de forma similar a como se hace con la transformada de Hough o mediante regresión lineal. *Sainz-Costa y col. (2011)*

desarrollaron una estrategia basada en el análisis de secuencias de vídeo para la identificación de las líneas de cultivo. Las líneas se mantienen en su posición considerando la proyección de perspectiva. Teniendo esto en cuenta, aplicaron la transformada de escala de grises basada en la aproximación de *Ribeiro y col. (2005)*, a continuación la imagen se binariza mediante umbralización basada en algoritmos genéticos. Cada imagen se divide en cuatro tiras horizontales. Se dibujan máscaras rectangulares sobre la imagen binaria para identificar las máscaras correspondientes a las filas y columnas. Los centros de gravedad de estas máscaras se usan como puntos de ajuste para definir las líneas de cultivo, seguidamente se ajusta una línea considerando estos puntos. La primera imagen de la secuencia se usa como una “look-up table” que sirve de guía en todo el proceso para determinar la posición de las siguientes máscaras en las sucesivas imágenes. *Hague y col. (2006)* transformaron la imagen original RGB a escala de grises. La imagen transformada se divide en ocho bandas horizontales. La intensidad de los píxeles a través de estas bandas muestra una variación periódica, debido a que las líneas de cultivo son paralelas. Puesto que las características de la cámara, su ubicación y la separación de las líneas de cultivo son conocidas a priori, la separación en píxeles de la imagen puede ser calculada para cada banda horizontal usando el modelo teórico de la cámara y su óptica asociada, dicho modelo teórico se conoce técnicamente, en terminología inglesa, como “pinhole” (*Tsai, 1987*). Se puede utilizar un filtro paso banda para realizar esta separación con su correspondiente respuesta característica en el dominio de la frecuencia. Algunas rejillas son difíciles de extraer porque las líneas y los cultivos forman una única máscara.

### **2.2.2. Métodos basados en la transformada de Hough**

De acuerdo con *Slaughter y col. (2008)*, uno de los métodos más utilizados para la detección de líneas de cultivo está basado en la transformada de *Hough (1962)*. La idea está basada en las líneas discontinuas donde existen “calvas” debido a una mala germinación o a otros factores que finalmente provocan zonas sin germinar. Está ideado para actuar en tiempo real y servir para el guiado de vehículos (*Marchant, 1996; Hague y col., 1997; Astrand y Baerveldt, 2005; Leemans y Destain, 2006*). Se aplica sobre imágenes binarias obtenidas aplicando métodos de binarización ampliamente conocidos, por ejemplo la transformación de imágenes RGB a escala de grises y posterior binarización (*Tellaeche y col., 2008a-b, 2011*). *Gée y col. (2008)*, aplican una doble

transformada de Hough con la asunción de que las líneas de cultivo son las únicas líneas de la imagen que convergen en el punto de fuga, el resto de líneas son rechazadas, otras restricciones como el espaciado entre líneas y restricciones geométricas sobre la proyección de perspectiva ayudan a identificar las líneas. Es necesario determinar un umbral en la transformada de Hough para hallar los valores pico (*Jones y col. 2009a-b*) o picos predominantes (*Rovira-Mas y col. 2005*). Dependiendo de las densidades en las líneas, sería factible suponer la existencia de varias de ellas, de suerte que un proceso de fusión posterior podría aplicarse a líneas con parámetros similares (*Tellaeche y col. 2008a-b, 2011*). *Ji y Qi (2011)*, abordaron la problemática de aplicar la transformada de Hough debido a la enorme cantidad de cálculo requerido y el consecuente tiempo de cómputo necesitado, proponen la aplicación de la transformada de Hough de forma aleatoria para reducir el tiempo de cómputo. Se han propuesto algunas modificaciones para mejorar la transformada de Hough como por ejemplo la realizada por *Asif y col. (2010)*, el cual aplica la transformada de Hough sólo a los puntos de borde a lo largo del cultivo. Sin embargo esto requiere la aplicación de técnicas para la extracción de bordes, que podría incrementar el coste computacional. Siguiendo la misma idea *Ji y Qi (2011)* han propuesto una transformada de Hough aleatoria con el mismo objetivo. La idea es evitar cálculos redundantes en la transformada de Hough. La realización práctica supone un proceso iterativo tomando muestras aleatorias de conjuntos de puntos para calcular una localización única en el espacio de Hough. Dado que dos puntos son trivialmente colineales, los parámetros de la ecuación de la recta que los une se pueden estimar fácilmente. Estos parámetros se utilizan para incrementar la celda del acumulador en el espacio de Hough.

Resumiendo, la transformada de Hough es computacionalmente costosa, mientras que su versión aleatoria requiere que los pares de líneas seleccionados sean considerados como una única línea recta, es decir, pares de puntos pertenecientes a una fila de cultivo. Si aplicamos esta técnica en imágenes donde los puntos de borde se han extraído previamente la selección de esos pares de puntos se vuelve complicada. Igualmente, el coste computacional de los algoritmos basados en la mencionada transformada de Hough es muy sensible a la resolución de la imagen tras un submuestreo de la misma. Además, cuando aparecen malas hierbas distribuidas de forma irregular, como es nuestro caso, puede provocar que el algoritmo no sea válido, este efecto se incrementa cuantas más malas hierbas haya.

### 2.2.3. Métodos basados en el punto de fuga

*Pla y col. (1997)* propusieron identificar regiones (cultivos, malas hierbas y tierra) mediante técnicas exclusivas de segmentación. Usan un esqueleto de cada región definida como característica para trazar las líneas de cultivo. Los esqueletos resultantes de cada región pueden usarse como curvas que definen la estructura de la línea para extraer líneas rectas donde se ubican las plantas y la tierra. Se definen segmentos en los esqueletos en forma de cadenas de puntos de contorno conectados que deben tener una longitud determinada. Esto permite elegir rectas candidatas asociadas a líneas de cultivo entre todas las posibles como aquellas que convergen en el punto de fuga. El punto de fuga se detecta utilizando información previa sobre las imágenes iniciales permitiendo un sencillo seguimiento del mismo. Este proceso es enormemente dependiente de los esqueletos, que no siempre son fáciles de aislar, teniendo en cuenta que los cultivos y las malas hierbas suelen solaparse.

### 2.2.4. Métodos basados en regresión lineal

Algunos de los métodos mencionados anteriormente aplican estas aproximaciones basadas en regresión lineal. *Billingsley y Schoenfish (1995, 1997)*, idearon un sistema de detección que es relativamente insensible al ruido adicional provocado por las malas hierbas. Usaron la regresión lineal en cada uno de los tres segmentos de cultivo y una función de coste análoga al momento de la línea que mejor se ajusta para detectar las líneas de cultivo, es decir, ruido y malas hierbas, como medio para identificar las líneas. Como hemos mencionado previamente *Søgaard y Olsen (2003)*, también aplicaron regresión lineal, lo cual es factible siempre y cuando la densidad de malas hierbas no sea alta y los píxeles pertenecientes a las líneas de cultivo estén claramente separados. En caso contrario se ven muy afectadas por píxeles que pertenecen a malas hierbas debido a su fuerte contribución en la estimación de la línea.

### 2.2.5. Métodos basados en estereoscopía

*Kise y col. (2005)* y *Kise y Zhang (2008)* desarrollaron un sistema de navegación para seguir las líneas de cultivo basado en la visión estéreo. El procesado de las imágenes estéreo se utiliza para determinar ubicaciones en 3D de puntos de interés de la imagen. Esas posiciones 3D, obtenidas mediante disparidad estereoscópica,

proporcionan la información suficiente para crear un mapa de elevación que usa una matriz de intensidades variable para indicar la altura del cultivo respecto del suelo y las malas hierbas que le rodean. Esta aproximación requiere cultivos con cierta altura. Debido a que el tratamiento del maíz ha de llevarse a cabo en sus primeras semanas donde la altura no es tan significativa, el método propuesto no resulta efectivo en nuestro caso. *Rovira-Mas y col. (2003, 2008, 2010)* han aplicado con éxito técnicas estereoscópicas en otras áreas y cultivos dentro de la agricultura de precisión, siempre con el propósito de obtener información 3D de la escena tanto para guiado como para detección de elementos de interés. En general, los sistemas de detección de líneas de cultivo basados en la visión estereo son prácticos si los cultivos tienen una cierta altura respecto del suelo y a su vez las malas hierbas y los cultivos difieren en altura.

### **2.2.6. Métodos basados en análisis de gotas (burbujas)**

Este método busca y caracteriza regiones de píxeles contiguos del mismo valor en imágenes binarizadas (*Fontaine y Crowe 2006*). El algoritmo busca gotas blancas (espacios entre filas) de más de 200 píxeles dado que las gotas más pequeñas podrían deberse al ruido entre las líneas de cultivo. Una vez que las gotas han sido identificadas, el algoritmo determina el ángulo del eje principal y localiza el centro de gravedad de la gota. Para una línea recta blanca perfecta, el centro de gravedad de la gota se sitúa sobre el centro de la línea de la propia gota, y el ángulo representa el ángulo de los espacios entre líneas. El algoritmo devuelve el ángulo y el centro de gravedad de la gota más cercana al centro de la imagen. En imágenes con abundantes malas hierbas la identificación de las gotas se convierte en una tarea ciertamente complicada ya que el solapamiento de las malas hierbas con las líneas de cultivo producen gotas muy amplias y con poca precisión.

### **2.2.7. Métodos basados en la acumulación de plantas verdes**

*Olsen (1995)* propuso un método basado en el hecho de que a lo largo de una línea de cultivo se produce una gran acumulación de zonas verdes en la imagen. La imagen se transforma a escala de grises donde las partes verdes aparecen más claramente que el resto. A continuación se obtiene una suma acumulativa de niveles de grises para una zona rectangular determinada de la imagen mediante la exploración de todas las columnas del rectángulo. Se da por hecho que las líneas verticales siguen esta

dirección en la imagen. Esto es, debido a que las imágenes son captadas por una cámara en posición ortogonal, evitando así las restricciones propias de la proyección de perspectiva. Se ajusta una curva sinusoidal mediante mínimos cuadrados. Los máximos locales de la senoide proporcionan la ubicación de los centros de las líneas. Este es un método sencillo y práctico que puede incluso simplificarse aún más. Sin embargo no es de nuestro interés dado que las imágenes captadas por la cámara del tractor tienen una proyección de perspectiva y precisamente no resulta conveniente tratar de evitarla. En nuestro algoritmo también utilizamos la idea de la acumulación de píxeles verdes a lo largo de la línea de cultivo mediante una estrategia más sencilla y menos costosa computacionalmente.

### **2.2.8. Métodos basados en análisis de frecuencia**

Debido a que las líneas de cultivo son paralelas en la escena 3D, en la imagen aparecen proyectadas bajo la ya mencionada proyección de perspectiva. Este hecho conlleva un comportamiento característico en el dominio de la frecuencia. *Vioix y col. (2004)* hacen uso de ello aplicando un filtro de Gabor bidimensional definido como una modulación de una función Gaussiana por una señal cosenoidal. El parámetro de la frecuencia requerido por el filtro de Gabor se deduce empíricamente de la transformada de Fourier 2D (*Bossu y col. 2006*). *Bossu y col. (2006)* aplican *wavelets* para discriminar las líneas de cultivo basándose en el análisis de frecuencia. Estos autores explotan el hecho de que las líneas de cultivo están bien localizadas en el dominio de la frecuencia, por tanto seleccionan una *wavelet* madre con esta frecuencia y las líneas de cultivo pueden extraerse por concordancia con dicha frecuencia. En campos de maíz donde llevamos a cabo nuestros experimentos, las líneas de cultivo muestran un comportamiento poco característico en el espacio de Fourier y por ello la aplicación de filtros basados en frecuencias no es un método práctico para nuestro propósito.

## **2.3. Descripción del método**

### **2.3.1. Imágenes Utilizadas**

Las imágenes usadas para este estudio corresponden a cultivos de maíz. Fueron tomadas con una cámara Canon EOS 400D durante los meses de Abril y Mayo de 2011 en los campos experimentales de 1,7 Ha de La Poveda (Arganda del Rey, Madrid)

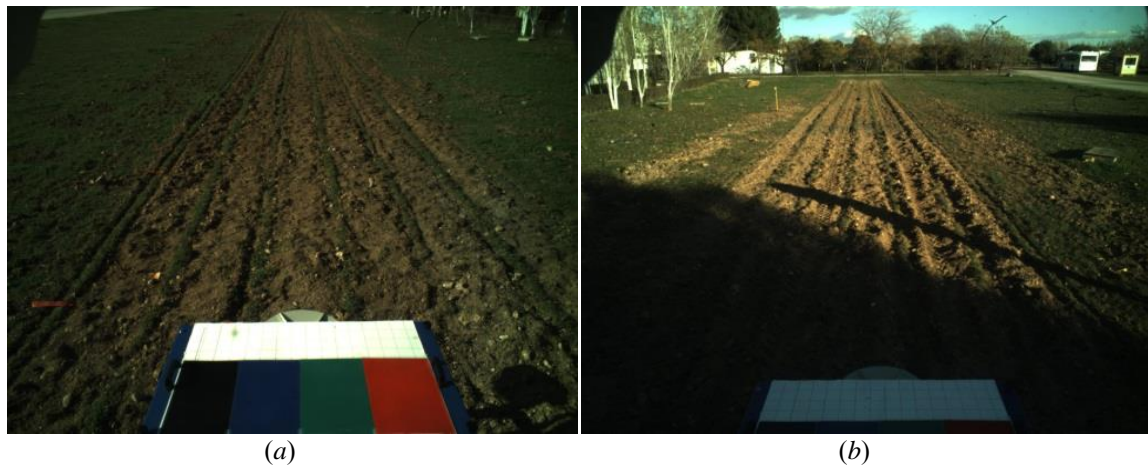
pertenecientes al Instituto de Ciencias Agrarias del CSIC (Consejo Superior de Investigaciones Científicas), las capturas se realizaron de forma que se espaciaron en el tiempo por más de cinco o seis días. A lo largo de ese periodo de tiempo se dieron diferentes condiciones atmosféricas que permitieron la creación de una base de datos de más de 350 imágenes. Toda esta variedad de imágenes nos permitió poner a prueba la robustez y eficacia del algoritmo ante situaciones lo más parecidas posibles a la realidad del trabajo relativo al procesamiento de imágenes en exteriores que constituye, como se ha indicado previamente, el objetivo del presente trabajo de investigación.

La variedad en las imágenes se consiguió de acuerdo con los siguientes criterios:

1. Diferentes condiciones de iluminación provocadas por las condiciones meteorológicas, *figura 2.4(a)-(b)*.
2. Diferentes condiciones de iluminación provocadas por la orientación del tractor respecto al sol, *figura 2.5(a)-(b)*.
3. Diferentes estados de crecimiento del cultivo y las malas hierbas, *figura 2.6(a)-(b)*.
4. Diferentes posiciones de la cámara, es decir, diferentes ángulos de inclinación, alabeo y orientación, *figura 2.7(a)-(b)*.
5. Diferentes densidades de malas hierbas, *figura 2.8(a)-(b)*.



**Figura 2.4:** Imágenes tomadas con diferentes niveles de iluminación debido a las condiciones atmosféricas del día: (a) nublado; (b) soleado.



**Figura 2.5:** Imágenes tomadas con diferentes orientaciones del sol: (a) tomada durante la mañana; (b) tomada al atardecer.



**Figura 2.6:** Imágenes tomadas en diferentes estados de crecimiento del cultivo: (a) Estado inicial (18 días después de la plantación); (b) Estado medio (33 días después de la plantación). Obsérvese en las ampliaciones el tamaño de la hoja de maíz debido a su estado de crecimiento.



**Figura 2.7:** Imágenes tomadas con diferente ángulo de orientación.



(a) (b)  
**Figura 2.8:** Imágenes con diferentes concentraciones de malas hierbas: (a) Densidad baja;  
 (b) Densidad alta.

Las imágenes fueron tomadas con las líneas de cultivo bajo proyección de perspectiva y almacenadas con color de 24-bits en el espacio RGB con resolución de 5 Mb y en formato JPEG.

El procesamiento de las imágenes inicialmente se llevó a cabo con Matlab R2009b (*The Mathworks 2014*) bajo Windows 7 con un procesador Intel Core 2 Duo T6600 a 2,4 GHz y 4GB de RAM, en etapas posteriores se han migrado los algoritmos a un procesador CompactRio de la compañía *National Instruments (2014)*.

### 2.3.2. Extracción de Píxeles Verdes de la Imagen

El algoritmo descrito en el presente capítulo se compone de dos etapas bien diferenciadas:

- a) Extracción de píxeles verdes de la imagen (también llamada binarización selectiva o segmentación).
- b) Detección de líneas de cultivo sobre la imagen binarizada.

La primera de ellas consiste en un procesado previo de la imagen necesario para eliminar elementos (píxeles) innecesarios para nuestro propósito que no harían otra cosa que interferir y falsear los resultados. Los píxeles eliminados son aquellos que no pertenecen a plantas, es decir, píxeles sin una componente espectral verde de suficiente peso, como pueden ser píxeles procedentes de la tierra del suelo, de piedras, ramas u otros elementos presentes en el campo de cultivo que no son vegetación. La segunda etapa comprende el algoritmo de detección de líneas propiamente dicho una vez que la

imagen ha quedado limpia de píxeles que no pertenecen a plantas. A continuación pasamos a describir la primera de las etapas, la extracción de píxeles verdes de la imagen.

En aplicaciones de tiempo real resulta de gran relevancia simplificar el proceso al máximo para poder cumplir con los tiempos de computación requeridos. Por ello, en lugar de usar los índices de vegetación (*Woebbecke y col., 1995; Ribeiro y col., 2005*), que requieren una transformación del espacio de color RGB al de escala de grises, utilizamos un método de aprendizaje cuya finalidad es la obtención de un umbral que permitirá identificar la componente espectral verde en el espacio de color mencionado, que consideramos como característica identificativa de las plantas en su estado de tratamiento, es decir con predominio del verde. Este método es necesario debido a la alta variabilidad de dicha componente como consecuencia de trabajar en exteriores ya que, las condiciones de iluminación se presentan continuamente cambiantes siendo muy difícil establecer un valor fijo inicial y tomarlo como constante durante toda la fase de operación del tractor a lo largo del tratamiento sobre el cultivo.

El procedimiento para la extracción de verdes se basa en el método de agrupamiento difuso que describimos a continuación:

De la imagen inicial extraemos aleatoriamente  $n$  muestras de entrenamiento, almacenándolas en el vector  $X$ , es decir,  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} \in \mathfrak{R}^d$  donde  $d$  es la dimensión de los datos (en nuestro caso  $d = 3$  por ser tres las componentes,  $r$ ,  $g$  y  $b$  en el modelo RGB de color utilizado), siendo  $\mathbf{x}_i$  las componentes espectrales RGB del píxel  $i$ .

Cada muestra se asigna a una clase  $w_j$ , donde el número de posibles clústeres es  $c$ , es decir,  $j = 1, 2, \dots, c$ . En nuestro caso  $c = 2$  porque sólo estamos interesados en dos tipos de texturas: las plantas verdes (cultivos, malas hierbas) y el resto (tierra, piedras, ramas secas, etc.).

Las muestras en  $X$  se clasifican utilizando el conocido método de agrupamiento difuso *FC* (*fuzzy clustering*, en terminología inglesa). El método toma las muestras  $\mathbf{x}_i$  y establece una partición, asumiendo que el número de clases es conocido. El problema de la partición difusa para  $b > 1$  es minimizar el criterio de varianza definido a continuación.

$$\min J_b(\mu_{ij}, \nu, t) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij}(t))^b \|\mathbf{x} - \mathbf{v}_j\|^2 \quad (2.1)$$

El proceso establece por cada  $\mathbf{x}_i$  en la iteración  $t$  el grado de pertenencia  $\mu_{ij}$  de dicha muestra a la clase  $w_j$ , y actualiza el centro de clases  $\mathbf{v}_j$ , de la forma expresada en la ecuación (2.1) (Zimmermann 1991).

$$\mu_i^j(t+1) = \frac{1}{\sum_{r=1}^c \left( \frac{d_{ij}(t)}{d_{ir}(t)} \right)^{\frac{2}{b-1}}} \quad (2.2)$$

$$\mathbf{v}_j(t+1) = \frac{\sum_{i=1}^n [\mu_i^j(t)]^b \mathbf{x}_i}{\sum_{i=1}^n [\mu_i^j(t)]^b} \quad (2.3)$$

$d_{ij}^2 \equiv d^2(\mathbf{x}_i, \mathbf{v}_j)$  es la distancia Euclídea. El número  $b$  es el peso exponencial (Bezdek, 1981; Duda et al., 2000),  $b > 1$ .

El criterio de parada del proceso iterativo se alcanza cuando se cumple la condición  $\|\mu_i^j(t+1) - \mu_i^j(t)\| < \varepsilon \forall ij$  o se alcanza un número máximo  $t_{max}$ , de iteraciones.

Este método requiere la inicialización de los centros de clases de tal forma que la ecuación (2.1) pueda aplicarse en la iteración  $t=1$ . Para ello, aplicamos el procedimiento descrito en Balasko y col. (2008):

1. Realizamos una transformada lineal  $Y=f(X)$  de las muestras de entrenamiento confinándolas al intervalo  $[0, 1]$ .
2. Inicializamos  $\mathbf{v} = 2D\bar{\mathbf{M}} \circ \mathbf{R} + D\bar{\mathbf{m}}$ , donde  $\bar{\mathbf{m}}$  es el vector medio de las muestras transformadas en  $Y$  y  $\bar{\mathbf{M}} = \max(\text{abs}(Y - \bar{\mathbf{m}}))$ , ambos con dimensiones  $I \times d$ ;  $D = [1 \dots 1]^T$  con dimensión  $c \times I$ ;  $R$  es una matriz de dimensión  $c \times d$  de números aleatorios con valores en el rango  $[0,1]$ ; la operación  $\circ$  denota multiplicación elemento por elemento.

Una vez que el proceso de aprendizaje ha finalizado obtenemos dos centros de clases  $v_1$  y  $v_2$  asociadas a las clases  $w_1$  y  $w_2$  respectivamente. Sin pérdida de generalidad, sea  $v_1 \equiv (v_{1R}, v_{1G}, v_{1B})$  el vector asociado a la clase que representa las plantas verdes. Es un vector tridimensional cuyas componentes,  $v_{1R}$ ,  $v_{1G}$  y  $v_{1B}$  representan los valores promedio  $R$ ,  $G$  y  $B$  de las componentes espectrales. Por tanto, el umbral para discriminar entre plantas verdes y el resto se establece finalmente en:

$$T_G = \frac{v_{1G}}{(v_{1R} + v_{1G} + v_{1B})} \quad (2.4)$$

Así pues, todo píxel de la imagen bajo procesamiento, cuyas componentes  $R$ ,  $G$  y  $B$  dadas por  $v_R$ ,  $v_G$  y  $v_B$  cumplan la relación expresada en la ecuación (2.3), se identificará como píxel verde perteneciente a una planta, mientras que aquellos otros que no cumplan la desigualdad serán considerados como tierra, roca u otros elementos distintos de vegetación ajenos a nuestro interés.

$$\frac{v_G}{v_R + v_G + v_B} > T_G \quad (2.5)$$

Una vez realizado este proceso obtendremos una imagen binaria donde los píxeles distintos del blanco pertenecen a plantas, bien de cultivo bien de malas hierbas, *figura 2.9*.



**Figura 2.9:** (a) Imagen original con líneas de cultivo; (b) imagen resultante binaria donde sólo permanecen los píxeles verdes pertenecientes a plantas de cultivo o malas hierbas (derecha).

### 2.3.3. Detección de Líneas

Una vez binarizada la imagen con los píxeles verdes identificados, sin ramas, piedras, tierra ni cualquier otro elemento distorsionador, estamos en disposición de aplicar el algoritmo para la detección de las líneas de cultivo.

Para ello, tenemos en cuenta las siguientes consideraciones, las cuales no afectan en absoluto a la validez de los resultados. El objetivo principal de dichas consideraciones es reducir el tiempo de cómputo, tanpreciado en procesos de tiempo real como es nuestro caso, pero sin pérdida alguna de robustez ni precisión en los resultados:

1. El número de líneas de cultivo a detectar es conocido de antemano. El conocimiento de este dato simplemente reduce el número de operaciones, dado que la búsqueda se restringe al número de líneas solicitado, es decir, matemáticamente hablando, acotamos el número de máximos locales a detectar en los valores obtenidos tras aplicar el algoritmo, como se verá posteriormente. Lo que se traduce en buscar en la gráfica que analizaremos más adelante (*figura 2.12*) tantos máximos locales como líneas de cultivo haya.
2. Las líneas de cultivo presentan proyección de perspectiva debido a la posición de la cámara en el tractor enfocando con un ángulo de inclinación en el sentido de la marcha. Por ello, dichas líneas en la imagen divergen en la parte inferior y convergen en la parte superior. Idealmente, si el tractor no estuviera sometido a vaivenes debido a la presencia de baches o irregularidades del terreno, la posición de las líneas en la imagen sería constante y el proceso para su detección sería relativamente fácil o incluso innecesario. Sin embargo, debido a la orografía del terreno, esa situación ideal es improbable y la posición de las líneas está en constante cambio respecto a la posición geoméricamente correcta. Estas variaciones serán tanto más acusadas cuánto mayores sean las irregularidades del terreno, las oscilaciones y las vibraciones de la cámara. Partiendo de este hecho, el algoritmo propuesto realiza una búsqueda de líneas de cultivo teniendo en cuenta la geometría ideal que deberían presentar las líneas aunque con ciertos márgenes de tolerancia. Así pues, consideramos como “línea de cultivo” toda línea que “nazca” en la parte inferior de la imagen, las líneas que nacen en los laterales izquierdo y derecho de la imagen no son consideradas de interés, *figura 2.10*. Esta restricción se aplica únicamente debido a la nula utilidad en nuestro proyecto que nos aportan las líneas que

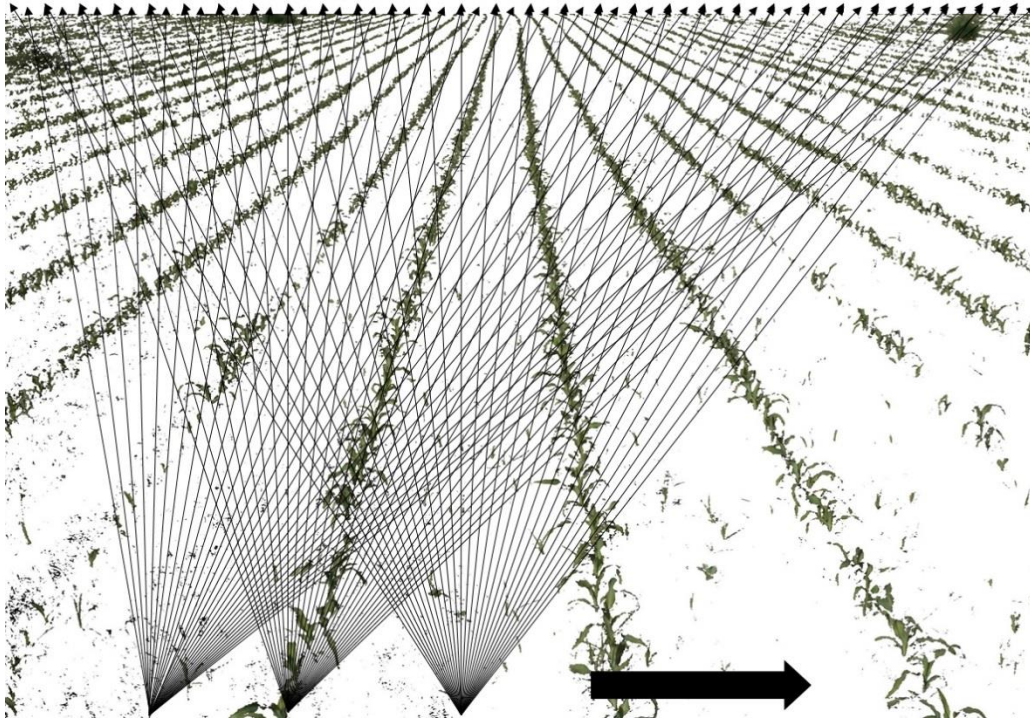
nacen en los laterales de la imagen. No obstante, este procedimiento sigue siendo válido para la detección de estas líneas descartadas, sin más que extender el procedimiento que veremos a continuación a los laterales de la imagen. Debemos aclarar, sin embargo, que esta ampliación conlleva el consiguiente incremento en el tiempo computacional.



**Figura 2.10:** Líneas de interés marcadas en negro, que nacen en la parte inferior de la imagen. Las líneas marcadas en rojo carecen de interés para nuestros propósitos y no son consideradas por el algoritmo.

A continuación describimos los pasos del método para detectar las líneas de cultivo:

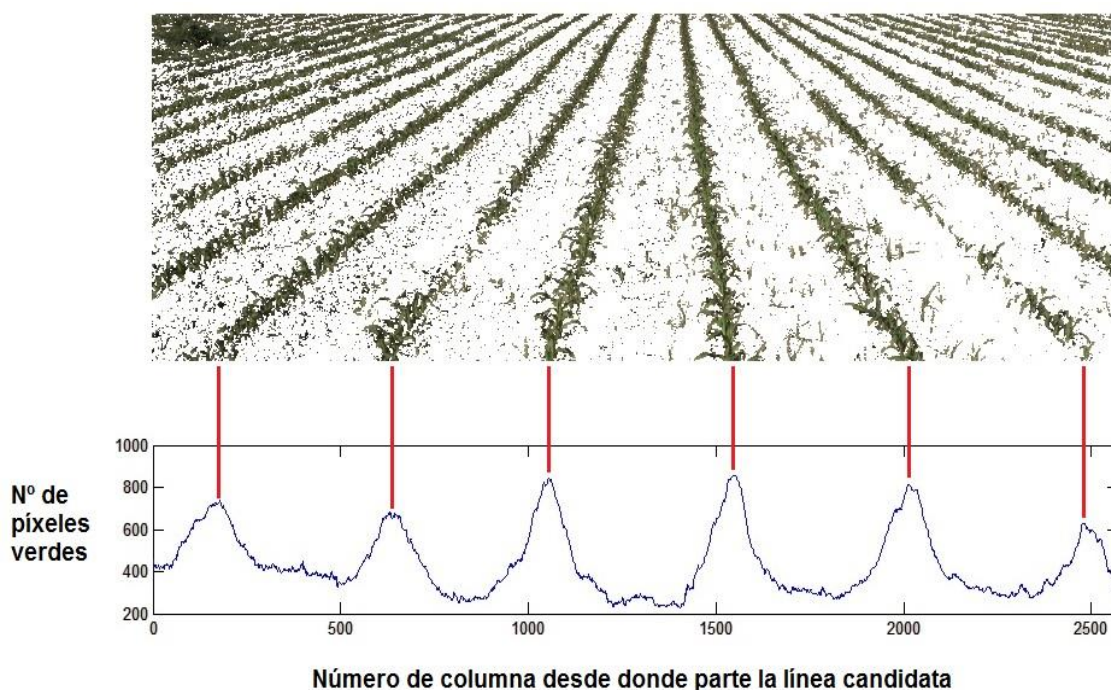
1. Desde cada píxel de la última fila de la imagen, trazamos una línea hasta cada píxel de la primera fila de la imagen tal y como se ilustra en la *figura 2.11*. Así pues, si la imagen tiene  $n$  columnas de resolución, tendremos que trazar  $n^2$  líneas. Como veremos más adelante este número de líneas a trazar es considerablemente reducido con las restricciones aplicadas en la *sección 2.3.3 “Detección de líneas”*.
2. Contamos el número de píxeles verdes que atraviesa cada una de las  $n^2$  líneas que hemos trazado en su camino desde abajo hasta arriba. Así pues, como de cada píxel de la fila inferior parten  $n$  líneas, tendremos  $n$  valores de píxeles verdes atravesados y asociados a cada píxel de la fila inferior.
3. De los  $n$  valores asociados a cada píxel nos quedamos con el valor máximo, es decir, elegimos de cada píxel de la fila inferior la línea con el mayor número de píxeles verdes atravesados. Esta línea es la línea candidata a ser una línea de cultivo ya que es la que más píxeles verdes atraviesa en su camino desde ese píxel de la fila inferior hasta la fila superior.



**Figura 2.11:** Desde todos y cada uno de los píxeles de la última fila (fila inferior) se traza una línea hasta todos y cada uno de los píxeles de la primera fila (fila superior).

4. Si dibujamos una gráfica con estos  $n$  valores máximos obtenemos una gráfica como la mostrada en la *figura 2.12.*, donde podemos observar cómo, a medida que nos aproximamos al surco real de la imagen, el valor máximo  $n$ , asociado a cada píxel va creciendo, siendo el máximo de entre los máximos cuando coincide con el píxel donde “nace” la verdadera línea de cultivo.
5. Dado que conocemos el número de líneas de cultivo a detectar  $c$ , el problema se reduce a la búsqueda de los  $c$  máximos locales de la gráfica de la *figura 2.12.* En la práctica, al estar la cámara fija en el tractor y ser todas las imágenes muy similares salvo por los movimientos o desviaciones imprevistas del tractor en su marcha, la búsqueda de los máximos locales se realiza acotando las zonas en un entorno próximo al teóricamente correcto. El valor teóricamente correcto es aquel donde debería “nacer” la línea de cultivo si el tractor estuviese perfectamente alineado. Este margen de tolerancia alrededor del píxel teóricamente correcto tendrá que ser mayor cuanto más abrupto sea el terreno para compensar las pequeñas desviaciones provocadas por las irregularidades del terreno o la existencia de baches,

rocas y demás. Así pues, el margen de tolerancia es un parámetro que queda libre y ajustable en función de la orografía del terreno.



*Figura 2.12: Líneas candidatas a ser líneas de cultivo. Los máximos de la gráfica se corresponden exactamente con las verdaderas líneas de cultivo.*

### 2.3.4. Restricciones

Con el fin de conseguir un algoritmo computacionalmente eficiente desde el punto de vista práctico, es decir, utilizable en aplicaciones de tiempo real, se realizan una serie de restricciones con las que ni se pierde precisión ni robustez. Dichas restricciones son las siguientes:

1. Dependiendo de la resolución de la imagen, no es necesario que el trazo de líneas sea píxel a píxel sino que se puede elegir un salto de tamaño  $s$ , y realizar el trazo de líneas cada  $s$  píxeles. Para imágenes con aproximadamente 2000 píxeles de resolución a lo ancho, un salto de  $s=10$  es más que suficiente para obtener unos resultados totalmente satisfactorios. Así pues, en lugar de trazar, desde cada uno de los 2000 píxeles de la última fila, 2000 líneas hasta la fila superior (4.000.000 de líneas en total), se trazarían desde 200 píxeles de la última fila, 200 líneas

hacia la fila superior, de esta forma trazaríamos 40.000 líneas, en lugar de 4.000.000.

2. Debido a la proyección de perspectiva de las imágenes, no es necesario trazar todas las líneas desde cada píxel inferior hacia cada píxel superior sino sólo aquellas líneas cuya pendiente se corresponde con lo esperado dentro de un margen lógico. Es decir, las líneas a la izquierda de la imagen tendrán una pendiente en sentido horario mientras que las líneas a la derecha la tendrán en sentido antihorario. Dependiendo del movimiento del tractor y por tanto del descuadre de la imagen, dicha reducción puede aplicarse entre un 50% y un 95% de las líneas de cada punto.
3. Igualmente, dependiendo de la resolución de la imagen, el conteo de píxeles verdes incluidos en cada línea puede hacerse no fila por fila sino eligiendo nuevamente un salto  $s$ . Por lo tanto, para las imágenes utilizadas en la investigación, con una resolución de aproximadamente 2.000 filas, un salto de  $s'=10$  es suficiente para obtener resultados satisfactorios.

Para ilustrar la utilidad de estas restricciones, en lo que a tiempo de cálculo se refiere, veamos el siguiente ejemplo: supongamos una imagen con una resolución de 1.500 filas x 2.500 columnas. Sin restricción alguna habría que trazar  $2.500 \times 2.500 = 6.250.000$  líneas y por cada línea habría que realizar 1.500 conteos de píxeles. En total habría que realizar  $1.500 \times 6.250.000 = 9.375.000.000$  operaciones. De acuerdo con las restricciones, si estableciéramos un salto de 10 píxeles para la realización del proceso descrito en el algoritmo (restricción 1), es decir, trazar una línea cada 10 píxeles, y además si trazásemos líneas en un entorno del píxel teórico donde debería coincidir la línea de un 10% (restricción 2) y finalmente contásemos los píxeles verdes cada 10 píxeles (restricción 3), el número final de operaciones sería  $(2.500/10) \times (2.500/100) \times (1.500/10) = 937.500$  operaciones. Es decir, el número de operaciones se reduce en un factor de 10.000, lo cual representa un número de operaciones más que aceptable en los procesadores actuales para poder realizar procesos en tiempo real.

## 2.4. Resultados

Durante la fase de entrenamiento se han procesado 240 imágenes capturadas a lo largo de los últimos cinco años en los campos experimentales de pruebas de la Poveda pertenecientes al Instituto de Ciencias Agrarias y la Centro de Automática y Robótica, ambos del CSIC en Arganda del Rey (Madrid), así como en explotaciones de maíz en la Ribera del Manzanares en Madrid. De dichas imágenes se han extraído aproximadamente 40.000 muestras para el aprendizaje. Cada muestra representa un píxel con sus correspondientes componentes espectrales R, G y B. Las imágenes fueron seleccionadas para cubrir un amplio rango de situaciones, a saber, diferentes condiciones meteorológicas, distintas concentraciones de malas hierbas, distintos encuadres y enfoques de la cámara con respecto al cultivo incluyendo desde dos hasta ocho líneas de cultivo y diferentes estados de crecimiento de las plantas de cultivo. Las *figuras 2.4. a 2.8.* muestran varios ejemplos ilustrativos de las situaciones mencionadas.

La *tabla 2.1* muestra los porcentajes espectrales tanto de las plantas verdes como del terreno.

	<b>Media de las componentes espectrales (0-255)</b>	<b>Porcentaje de la componente espectral mayor</b>
<b><math>v_1</math> (plantas verdes)</b>	{137,80 140,68 106,07}	37% (verde) representando las plantas y malas hierbas
<b><math>v_2</math> (tierra y otros componentes)</b>	{188,49 177,71 153,53}	36% (rojo) representando la tierra, rocas de los entresurcos.

*Tabla 2.1. Porcentajes para cultivos (con color verde) y tierra (con color marrón).*

La extracción de verdes se ha realizado antes de ejecutar el algoritmo de detección de líneas de cultivo, para eliminar tierra y otros elementos distintos de la vegetación como se ha comentado previamente en la *sección 2.3.2 “Extracción de píxeles verdes de la imagen”*.

A efectos de comparar tiempos de cómputo, se ha implementado el algoritmo de Hough adaptado para la detección de líneas de cultivo con la disposición de las imágenes utilizadas. En la implementación del algoritmo de Hough se ha aplicado el mayor número de restricciones posibles para poder equiparar algoritmos similares teniendo en cuenta la geometría característica de las imágenes y con el máximo número de restricciones posibles sin que afecten a los resultados. Así pues, las líneas no

candidatas a ser líneas de cultivo no se han considerado en el algoritmo de Hough. Las restricciones realizadas han sido las siguientes:

1. El rango de ángulos de las líneas calculadas se ha limitado lo más cercano posible a los ángulos de las líneas de cultivo en lugar de calcular las líneas en los 360 grados posibles, como habría hecho el algoritmo de Hough sin restricciones, es decir, líneas horizontales y próximas a éstas se han ignorado.
2. Se establece un salto discreto, entre todos los posibles ángulos, de un grado.

La *tabla 2.2* y la *figura 2.17* muestran los promedios del tiempo de computación obtenido con el algoritmo de Hough (sin obtención del esqueleto) y nuestro algoritmo. Dichos cálculos se han realizado con los algoritmos implementados en Matlab R2009b (*The Mathworks 2014*), lo cual no da una idea del tiempo óptimo de cada algoritmo y su diferencia absoluta pero sí proporciona una comparación relativa entre ambos.

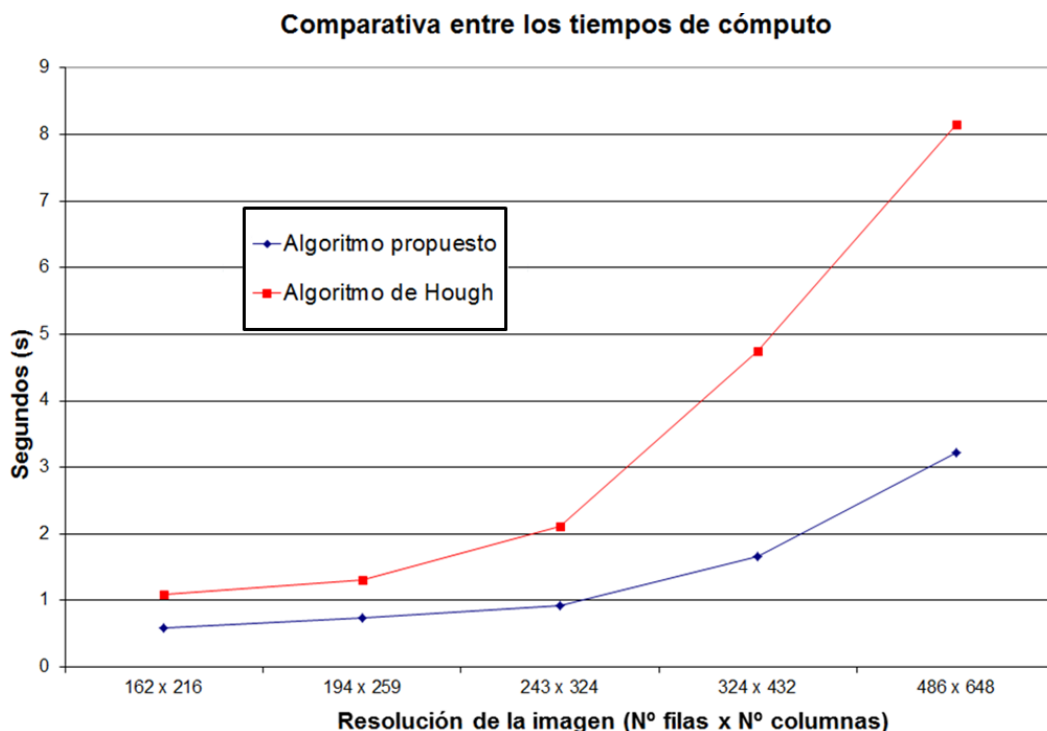
<b>Resolución de la imagen (píxeles)</b>	<b>Tiempo con algoritmo de Hough (segundos)</b>	<b>Tiempo con nuestro algoritmo (segundos)</b>	<b>Diferencia de tiempos (segundos)</b>	<b>% de diferencia de tiempos</b>
162 x 216	1,088	0,58	0,508	87,59
194 x 259	1,305	0,737	0,568	77,07
243 x 324	2,12	0,928	1,192	128,45
324 x 432	4,752	1,667	3,085	185,06
486 x 648	8,153	3,216	4,937	153,51

*Tabla 2.2: Promedios de tiempos de computación obtenidos con el algoritmo de Hough y nuestro algoritmo.*

A efectos de utilizar la transformada de Hough para la detección de líneas de cultivo habría que tener en cuenta no sólo el mayor tiempo de cómputo de éste respecto a nuestro algoritmo, sino también los siguientes inconvenientes:

1. Las malas hierbas pueden hacer errar el método de la transformada de Hough a la hora de determinar la línea de cultivo. Esto es debido a la proximidad de valores altos que se produce en el acumulador a causa de las malas hierbas. Así pues, cuanta mayor sea la concentración de malas hierbas, menor será su fiabilidad.

- Para evitar seleccionar líneas cercanas entre sí en el acumulador definimos “un parche” que colocamos alrededor de cada línea elegida de tal forma que una vez elegida una línea del acumulador y colocado el parche, no se elegirá ninguna línea próxima a la línea elegida. La elección del tamaño de este parche hay que hacerla de forma aproximada e intuitiva. Esto provoca errores en numerosas ocasiones por la elección de un parche excesivamente grande o pequeño según el caso.



*Figura 2.17: Promedios de tiempos de computación obtenidos con el algoritmo de Hough y nuestro algoritmo en función de la resolución de la imagen.*

Por tanto, nuestro método no sólo es más rápido que el algoritmo de Hough sino que además es más robusto y fiable a la hora de detectar líneas de cultivo en imágenes con proyección de perspectiva como las tomadas por la cámara del tractor.



## CÁLCULO DE LAS COORDENADAS X-Y DE OBJETOS EN IMÁGENES

### 3.1. Introducción

El presente algoritmo ha sido desarrollado como parte del proyecto perteneciente al Plan Nacional de I+D+i titulado “*Sistema de Vigilancia, Búsqueda y Rescate en el Mar mediante colaboración de Vehículos Autónomos Marinos y Aéreos*” (DPI2009-14552-C02-01) con el objetivo de determinar las coordenadas de un barco no tripulado sobre una superficie de agua como puede ser el mar, un lago o un pantano, con la única ayuda de una cámara posicionada en tierra. Calcula las coordenadas espaciales  $x$ - $y$  de cualquier objeto sobre dicha superficie y presente en una imagen respecto a un origen de coordenadas previamente establecido, además proporciona la distancia lineal del objeto a la cámara en ambos casos con respecto al origen de coordenadas elegido.

Igualmente, dicho algoritmo es utilizado en el proyecto *RHEA (2014)* para ubicar mediante medidas de posición y distancias, las zonas de malas hierbas que aparecen en las imágenes proporcionadas por el tractor. En este caso el origen de coordenadas está ubicado en la propia cámara. Dado que el tractor está provisto de un sistema GPS (Global Position System) y por tanto su posición real es conocida, con este algoritmo podemos conocer la ubicación relativa de las zonas de malas hierbas con respecto a la cámara y por tanto, mediante un cambio de coordenadas cualquier elemento en la imagen situado sobre el suelo, incluyendo las malas hierbas, queda perfectamente localizado. No se trata de un algoritmo de propósito general sino que requiere de dos condiciones específicas necesarias para poder ser aplicado, a saber:

1. Los objetos deben estar ubicados en superficies planas y con un alto grado de horizontalidad, por ejemplo, mar, campos llanos, suelos de edificios, explanadas, etc.
2. La cámara debe estar fija e inmóvil una vez realizados los ajustes descritos en la *sección 3.3.1 "Cálculo de la coordenada-y"*.

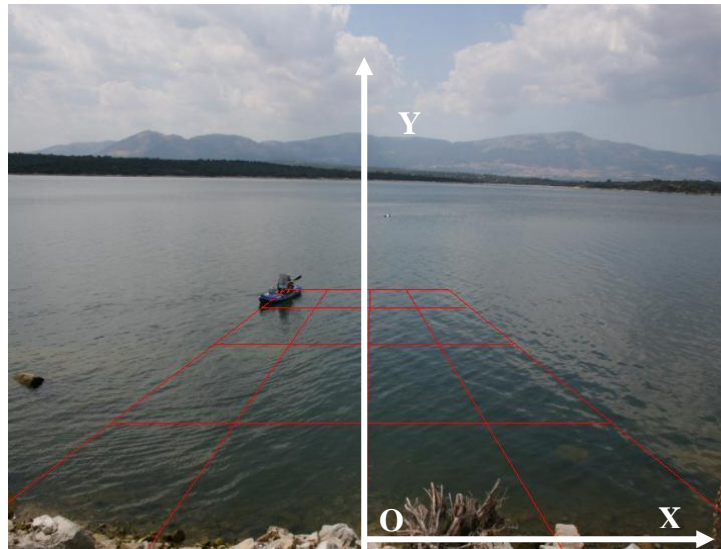
Tanto en el caso de los barcos no tripulados como en el de la localización en el campo se cumplen ambas condiciones, razón por la cual es utilizado en ambos proyectos.

Para la descripción del algoritmo vamos a utilizar las imágenes y el método seguido en el caso de la localización de barcos. No obstante, el procedimiento para la detección de malas hierbas es idéntico en ambos casos.

Por simplificación de cálculos en el método establecemos el origen de coordenadas de tal forma que coincida con el centro de la imagen y el *eje-X* sea paralelo a las filas de la propia imagen, mientras que el *eje-Y* se sitúa perpendicularmente al *eje-X* y en la dirección del eje óptico. En la *figura 3.1* se observa el posicionamiento del *eje-Y* coincidiendo el eje óptico de la cámara, y el centro *O* coincidiendo con el centro del sensor CCD (Charge-Coupled Device, dispositivo de carga acoplada) de la misma. En la imagen de la *figura 3.2* se muestra la imagen capturada por la cámara con la geometría descrita previamente.



**Figura 3.1:** Vista cenital de la escena con el eje-Y coincidente con el eje óptico de la cámara y el eje-X perpendicular al anterior.

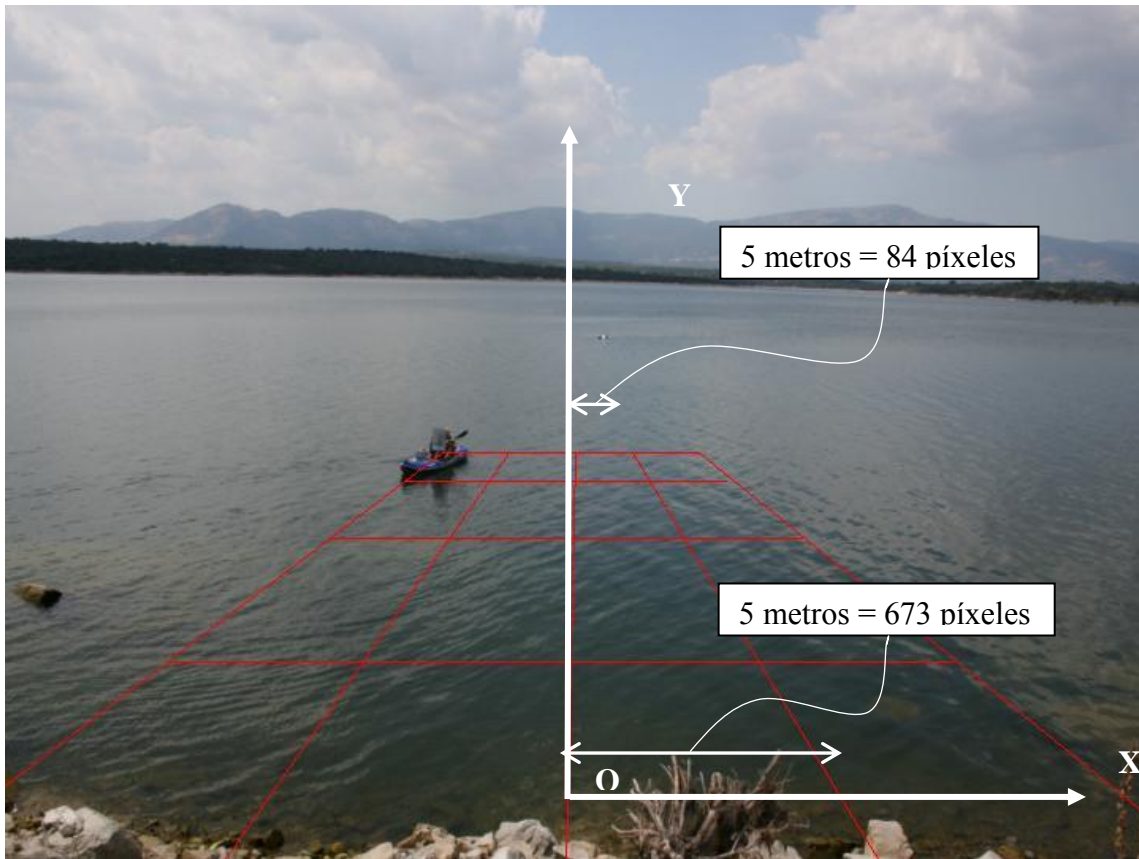


**Figura 3.2:** Vista en perspectiva de la misma escena de la figura 3.1 tal cual se ve en la imagen capturada por la cámara.

El algoritmo se basa en dos propiedades inherentes a la proyección de perspectiva proporcionada por las cámaras convencionales, a saber:

1. Cuanto más próximo se sitúa un objeto en relación al observador más próximo aparece también respecto de la línea inferior de la imagen y viceversa, cuanto más alejado está el objeto más alejado de la línea inferior aparece en la imagen, *figura 3.5(a)*. En otras palabras, existe una clara relación entre la fila en la que aparece un objeto y la distancia al observador. Esta relación se puede modelar como una función matemática que relaciona la fila  $f$ , en la que aparece el objeto en la imagen, y la *coordenada-y* del objeto respecto al origen de coordenadas que hayamos elegido.
2. Cuanto más próximo esté el objeto al observador, mayor número de píxeles abarca sobre la imagen. Es decir, el tamaño de un objeto en una imagen no sólo depende de las dimensiones propias del mismo, sino también de la distancia a la que se encuentre del observador. Por ejemplo, un objeto situado a 5 metros del *eje-Y*, es decir,  $x=5m$ , tendrá una distancia en píxeles al *eje-Y* distinta dependiendo de en qué fila de la imagen se haya proyectado, *figura 3.3*. De nuevo, esto implica que la *coordenada-x* de un objeto depende obviamente de su distancia al *eje-Y* pero también de la *coordenada-y* en la que se encuentre. Nuevamente, esta otra relación se

modela matemáticamente y nos proporciona la dependencia de la *coordenada-x* con respecto a la fila  $f$ , y la columna  $c$ , en la que aparece el objeto en la imagen.

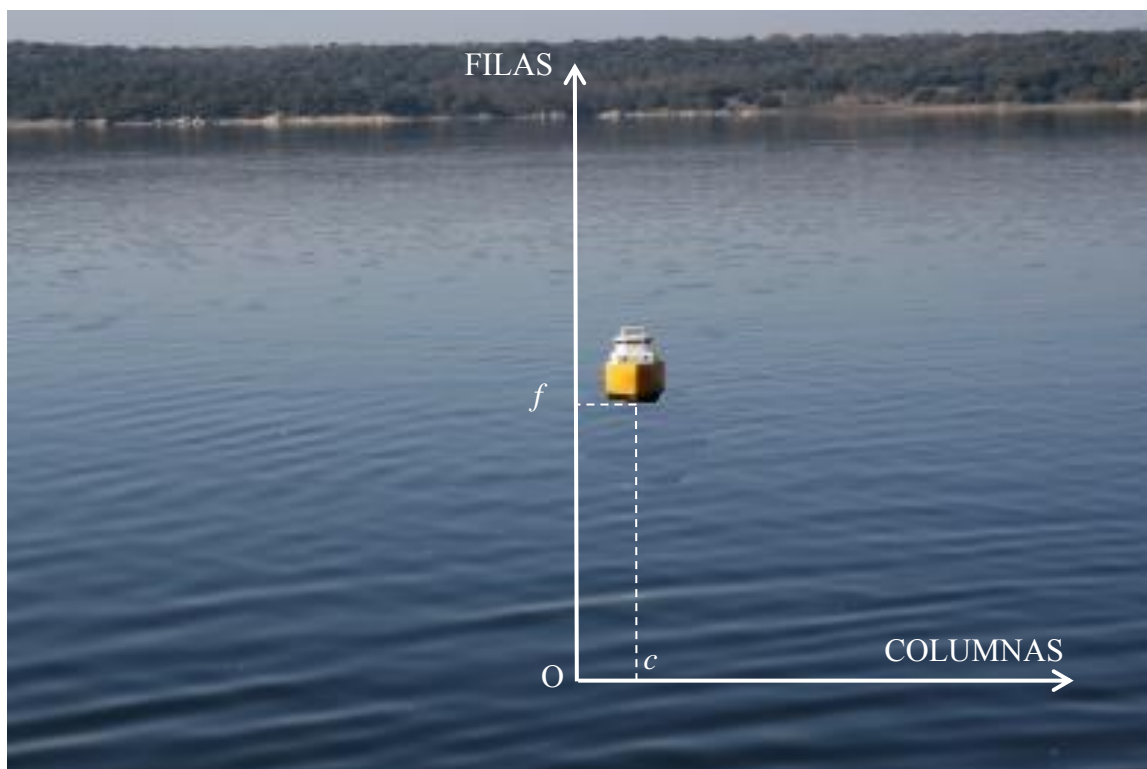


**Figura 3.3:** un punto a una misma distancia en el eje-X (5 metros) puede corresponder a más o menos píxeles dependiendo de su posición en el eje-Y.

Teniendo en cuenta estos dos hechos, el problema se reduce a la búsqueda de las funciones matemáticas mencionadas anteriormente, las cuales relacionan las *coordenadas x-y*, con la fila  $f$ , y la columna  $c$ , donde aparece en la imagen el objeto. Conocidas las funciones matemáticas, las *coordenadas x-y* de cualquier píxel de la imagen pueden ser calculadas sin más que sustituir en la fórmula la fila  $f$  y la columna  $c$  en que se encuentran. Dicha función se modela a partir de datos experimentales extraídos de antemano en el propio terreno a modo de calibración. Una vez obtenidos estos datos, la cámara ha de permanecer inmóvil para que tengan validez las funciones matemáticas calculadas en base a esos datos. Si la cámara se mueve es necesario volver a tomar datos experimentales y calcular las nuevas funciones matemáticas.

El presente método determina las coordenadas de un objeto a partir de su posición en una imagen. Para ello, como hemos visto, establecemos una relación

matemática entre las *coordenadas x-y* en relación a la fila  $f$  y columna  $c$  donde se encuentra el objeto. Hay que tener en cuenta que un objeto en una imagen ocupa muchos píxeles, no sólo uno. Por ello, a cada objeto del que queremos calcular sus coordenadas debemos asignarle un único píxel, es decir, una única fila y una única columna. Para ello, se toma el píxel más bajo que pertenezca al objeto y esté en contacto con la superficie, tal y como se muestra en la *figura 3.4*. Este píxel coincide con el punto de la superficie plana de contacto donde se encuentra realmente posicionado el objeto.



*Figura 3.4: Elección del píxel de referencia que corresponde a un punto del objeto en contacto con la superficie plana.*

## 3.2. Estado del arte

En los últimos años se han propuesto distintos métodos para la estimación de la profundidad de objetos. El campo que estudia dichos algoritmos se llama visión por computador 3D. Su principal objetivo es determinar la posición de objetos en una escena tridimensional. Los métodos existentes pueden clasificarse en dos categorías: métodos pasivos y métodos activos (*Kumar-Mada y col. 2002*). Los métodos activos aplican energía mediante una fuente estructurada (laser, radar, haces de luces) y reciben respuestas a través de un detector. Basándose en la naturaleza de la interacción con las

superficies de los objetos, la energía reflejada se usa para reconstruir la estructura 3D de la escena. De todos son conocidas las limitaciones en entornos no estructurados como los escenarios al aire libre donde la reflectancia de los objetos, los rangos de colores y la intensidad de iluminación son variables con un cambio continuo en sus valores. Por ello las técnicas activas no son muy recomendables en estos entornos. Sin embargo las técnicas pasivas no aplican fuentes de energía estructurada para reconstruir las imágenes evitando con ello interferencias con el entorno. El presente algoritmo está encuadrado dentro de los métodos pasivos. La mayoría de los métodos para la determinación 3D de una escena son métodos pasivos y se agrupan bajo la expresión genérica de “formas a partir de X”, donde X puede ser cualquiera de los conceptos que se relacionan a continuación (*Trucco y Verri 1998*):

### 3.2.1. Estereopsis

*Scharstein y Szeliski (2002)*, basándose en dos cámaras separadas una cierta distancia entre ellas, llamada línea base, un punto 3D en la escena es proyectado en diferentes coordenadas de filas  $r$ , y columnas  $c$ , en cada imagen. Se define un procedimiento de proyección para identificar el punto en ambas imágenes. Una vez que la correspondencia se ha establecido, la posición 3D del objeto se calcula mediante triangulación. El emparejamiento de píxeles (entre dos imágenes que cumplen la restricción epipolar, por tanto estéreo-relacionadas) es un tema crítico en visión estereoscópica que se convierte en inútil cuando existen grandes áreas homogéneas en la escena. Este es nuestro caso, dado que las superficies acuáticas son grandes extensiones de texturas más o menos homogéneas. A partir de ciertas distancias los barcos se proyectan como pequeños parches sin forma clara ni definida, haciendo muy difícil su identificación mediante técnicas de reconocimiento basadas en patrones. Además, el movimiento continuo de los barcos en la superficie del mar requiere que la receptividad a la luminosidad entre ambas cámaras tenga que ser muy parecida. La estimación de la profundidad se realiza normalmente utilizando técnicas basadas en el concepto de visión estereoscópica (*Scharstein y Szeliski 2002; Middlebury 2013*). En los últimos años se han introducido numerosas modificaciones de los métodos clásicos para determinar características de objetos en imágenes, por ejemplo, altura, volumen, distancia. En *Chen y col. (2006)*, se propone un método robusto para medir la altura del objeto basado en dos cámaras. Este método calcula la altura usando diferentes planos de la imagen y

realizando transformaciones geométricas. Por otro lado, existen algoritmos que sólo obtienen profundidades relativas y mapas de disparidad sin la localización de objetos en la escena (*Aziz y Mertsching 2010; Daolei y Lim 2011; Agrawal y col. 2007*). Otras implementaciones combinan algoritmos con el objeto de reducir el tiempo de computación en la obtención de mapas de profundidad (*Shan-Shan y col. 2011; Birchfield y Tomasi 1998*). La mayoría de estos métodos proporcionan información sobre la profundidad y características de los objetos en tiempos de computación aceptables. Sin embargo necesitan los parámetros intrínsecos y extrínsecos de ambas cámaras, no proporcionan información numérica (cuantitativa) sobre la distancia entre el objeto y las cámaras y además sólo han sido probados con resultados eficientes en interiores con las condiciones de iluminación suficientemente controladas. Existen también técnicas basadas en visión estereoscópica con resultados aceptables en entornos de exterior tales como agricultura con posicionamiento relativo entre los objetos de la escena y el sistema de visión (*Rovira-Mas y col. 2008*).

### **3.2.2. Movimiento**

Donde el objeto puede estar moviéndose y la cámara está estática (*Buch y col. 2009; Toshev y col. 2009*), o viceversa (*Leibe y col. 2007*). En imágenes con objetos en movimiento, las imágenes consecutivas producen áreas que cambian con respecto al fondo. Estos métodos calculan el movimiento relativo entre la cámara y la escena 3D con pequeños cambios. Estos cambios son detectados mediante análisis diferencial tanto en la dimensión espacial como en la temporal (flujo óptico). La reconstrucción 3D es una tarea compleja ya que la diferenciación requiere movimientos pequeños con unas pocas características en los objetos (*Ozden y col. 2010*). En nuestro caso los barcos están en continuo movimiento provocando pequeños, y no tan pequeños, desplazamientos.

Una forma de adquirir datos 3D es determinar la dirección de traslación a través de lo que se conoce como paralaje de movimiento aproximado. Las profundidades pueden ser obtenidas mediante estimaciones de mínimos cuadrados a partir del flujo óptico *Cantzler (2011)*. Debido al constante movimiento de los barcos ya mencionado, la reconstrucción 3D mediante el movimiento se ve fuertemente afectada y por ello resulta un método muy poco efectivo para la localización de barcos en superficies acuáticas y por tanto en entornos de exteriores.

### 3.2.3. Tonalidades

*Horn (1986), Zhang y col. (1999) y Prados y Faugeras (2006)* aplican patrones de luces y sus correspondientes sombras para deducir las formas. Se requiere conocimiento previo de los ángulos de incidencia de la luz y también una estimación de la forma de la superficie. Los análisis con sombras fallan cuando se producen cambios en la reflectancia, lo cual es muy común en entornos de exteriores. En nuestro caso este método es inviable dada la característica de nuestro entorno de exterior y que estamos en superficies planas. Los objetos en la superficie del agua, a ciertas distancias, no muestran suficientes sombras para poder realizar una reconstrucción 3D y tampoco muestran superficies conocidas. Además con estos métodos se obtienen profundidades relativas sin ningún sistema de referencia.

### 3.2.4. Texturas

Son métodos basados en la repetición de objetos o patrones dentro de la escena (*Payet y Todorovic, 2011; Lobay y Forsyth, 2006*). Los métodos basados en análisis de texturas hacen uso de la existencia de repeticiones regulares de un elemento o patrón llamado “texel” en la superficie de la escena. De este modo, fijándonos en estos texels, una vez que son proyectados sobre la imagen, la estructura 3D puede deducirse mediante el análisis de la distorsión de los mismos. Dicho análisis se realiza comparándolos mutuamente o calculando la distorsión producida debido a la proyección de perspectiva; por ejemplo, circunferencias en la escena se manifiestan como elipses una vez son proyectadas sobre la imagen. Esta última suposición requiere que las características de la textura sean constantes dentro de regiones de textura uniforme. En nuestro caso, esta repetición de objetos o de características no aparece en los escenarios de trabajo.

Igualmente, la superficie acuática es grande con tan sólo unos pocos barcos sobre ella. Por esta razón, los métodos basados en análisis de texturas no son prácticos para el caso de estudio en el que basamos la investigación. Además en estos métodos son necesarios conocimientos de los parámetros intrínsecos de la cámara, algo que requiere de un proceso de calibración bastante preciso para evitar al máximo los errores posteriores

### 3.2.5. Enfoque

Basado en que imágenes bien enfocadas se proyectan sobre el plano de la imagen con un alto grado de nitidez (*Krotkov, 1987; Subbarao y col., 1993; Subbarao y Surya, 1994*). Así pues, si conocemos la focal del sistema óptico, mediante un procedimiento de triangulación podemos determinar la posición 3D de los objetos en la escena. Estos métodos requieren ajustar dicha focal para determinar áreas bien enfocadas. Este hecho en superficies homogéneas no resulta nada práctico. Igualmente requiere un sistema motorizado para ajustar la lente en los enfoques, además de un calibrado de precisión (*Tao y col. 2010*). Existen distintas técnicas para determinar el grado del enfoque de un elemento en la imagen, la mayoría de ellas basadas en análisis de bordes, que determinan la nitidez de los mismos, por ejemplo calculando la varianza alrededor de un punto de borde, de forma que en bordes nítidos ésta debe ser alta frente a bordes desenfocados que es relativamente baja. En nuestro caso la superficie del agua no presenta bordes ni existen elementos que nos pudiesen servir para el enfoque a través de sus bordes. Además de que los bordes de los barcos a cierta distancia no poseen en la imagen la precisión requerida para aplicar métodos de enfoque.

### 3.2.6. Determinación de la profundidad con una sola cámara

*Saxena y col. (2008)* aplican aprendizaje supervisado, *Kaushik (2012)* determina correspondencias entre las distancias de los objetos y sus píxeles en la imagen. Aquí se engloban distintos métodos, así *Saxena y col. (2008)*, aplican aprendizaje supervisado donde las relaciones se establecen entre estructuras en la imagen y profundidad de la escena basadas en la disponibilidad de una referencia o “ground-truth”. Estas relaciones son aplicadas más tarde a nuevas imágenes en escenarios similares. Nuestro escenario, la superficie del agua, se muestra como una gran estructura con infinidad de niveles de profundidad. Esto hace inviable la aplicación de estos métodos.

*Wang y col. (2005)* y *Criminisi y col. (2000)*, consideraron puntos de fuga en la imagen y un plano de referencia en la escena para calcular la matriz de proyección de donde se derivan las alturas de los objetos. De la misma manera *Fernandes y Neves (2006)*, propusieron un método capaz de estimar distancias con una única cámara mediante una calibración específica utilizando identidades trigonométricas. Estos métodos están basados en un conocimiento previo de las alturas de los objetos reales.

*Hsu y col. (2009)* calculan profundidades basándose en desplazamientos de la cámara. Establecen una relación entre el desplazamiento de la cámara y la variación en la imagen de los píxeles del objeto en consideración. En los métodos expuestos anteriormente los objetos permanecen inmóviles y a relativamente cortas distancias. Todo lo contrario que en nuestro caso donde los barcos no permanecen inmóviles y se encuentran a grandes distancias.

*Kaushik (2012)* obtiene muestras de objetos de referencia de diferentes imágenes. Las posiciones de los objetos de referencia son conocidas en la escena 3D y también la *coordenada-y* del píxel en las imágenes. Calcula un polinomio de grado- $n$  que describe la *coordenada-y* en la escena 3D como una función de píxeles donde los objetos son localizados en la imagen.

*Salvi y col. (2002)* realizan una revisión sobre métodos de calibración de cámaras, entre los que se encuentran las técnicas lineales que obtienen la matriz de transformación directa, mediante mínimos cuadrados, entre las coordenadas en píxeles de la imagen y las coordenadas espaciales en unidades de longitud (*Hall y col., 1982; Faugeras y Toscani, 1986; Ito, 1991*). Debido a que las distorsiones radiales de las lentes son no lineales, estos métodos a veces no son lo precisos que cabría esperar.

El objetivo del algoritmo que se describe en este capítulo es desarrollar un método válido para posicionar barcos en una superficie acuática a distancias relativamente grandes (hasta 150 metros aproximadamente). Desde el punto de vista del análisis de texturas, las superficies acuáticas presentan una enorme homogeneidad, como ya se ha mencionado anteriormente. Debido a nuestras necesidades y a las limitaciones de los distintos métodos expuestos en esta sección, presentamos un algoritmo que es independiente de los parámetros intrínsecos y extrínsecos de la cámara, de tal forma que nos evitamos los problemas de calibración en un escenario tan complicado. La mayoría de los métodos presentados anteriormente proporcionan mapas de profundidad relativa en relación a otros objetos pero no dan las *coordenadas x-y* de un objeto con respecto a un origen de coordenadas. El método más próximo es el propuesto por *Kaushik (2012)*, aunque insuficiente ya que únicamente proporciona la *coordenada-y* del objeto. Al igual que *Kaushik (2012)*, nosotros también calculamos funciones matemáticas para establecer las relaciones entre las coordenadas reales del objeto y las filas y columnas en las que aparece en la imagen. Estas funciones, que en

nuestro caso no son polinomios, tienen suficiente precisión para nuestros propósitos. La principal aportación del método propuesto es que no sólo proporciona la distancia del objeto al observador sino que además calcula sus *coordenadas x-y* con respecto a un origen de coordenadas previamente establecido. Más concretamente la principal aportación de la investigación es el cálculo de la *coordenada-x* a partir de la fila  $f$ , y de la columna  $c$ , en la que aparece el objeto en la imagen. Además, como ya se ha mencionado, este método no requiere de calibración alguna ni de conocimientos sobre la geometría de la cámara o de sus parámetros intrínsecos. Finalmente, añadir que el método proporciona un posicionamiento de objetos respecto a un origen de coordenadas como lo hace un GPS, el cual es utilizado para comparar resultados.

### 3.3. Descripción del algoritmo

Como hemos visto en la *sección 3.1 “Introducción”* el problema consiste en el cálculo de dos funciones, a saber, una que relaciona la *coordenada-y* del objeto con la fila  $f$  en la que el objeto se encuentra en la imagen y otra que relaciona la *coordenada-x* con la fila  $f$ , y la columna  $c$ , en la que se encuentra el objeto dentro de la imagen.

#### 3.3.1. Cálculo de la *coordenada-y*

Como se ha mencionado previamente, la *coordenada-y* del píxel sólo depende de la fila en la que aparezca en la imagen, ya que cuanto más próximo al observador está el píxel más abajo aparecerá en la imagen. En la *figura 3.5(a)* podemos ver que para cualesquiera dos objetos con sus coordenadas respectivas  $y_1$  e  $y_2$ , y con filas en la imagen  $f_1$  y  $f_2$  se cumple la siguiente relación,

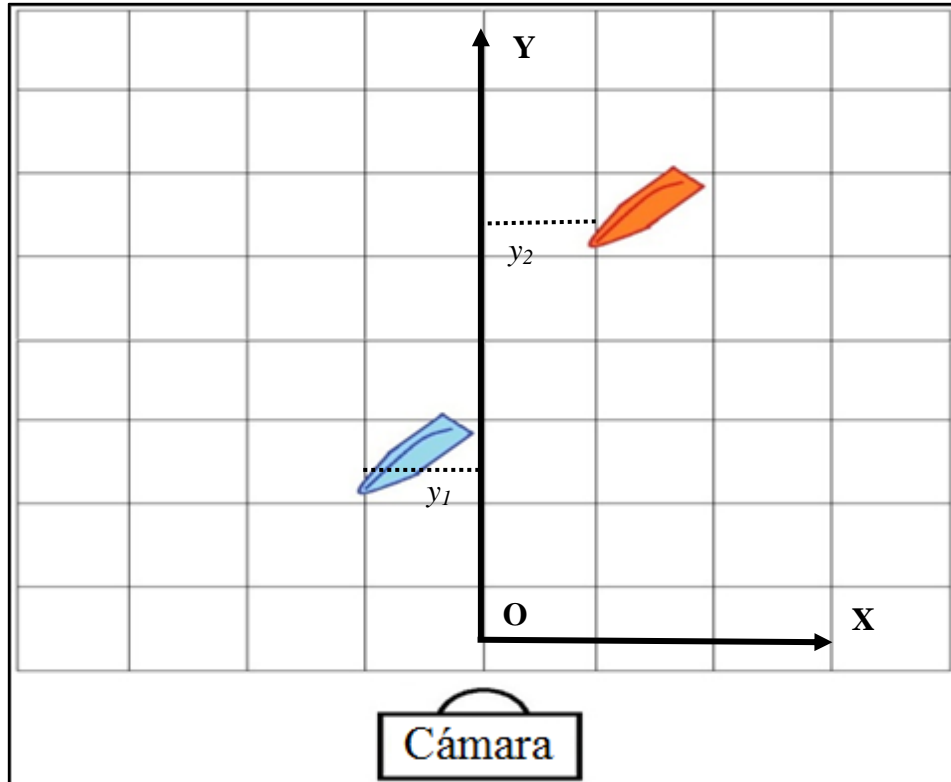
$$y_1 < y_2 \iff f_1 < f_2 \quad (3.1)$$

#### 3.3.2. Cálculo de la *coordenada-x*

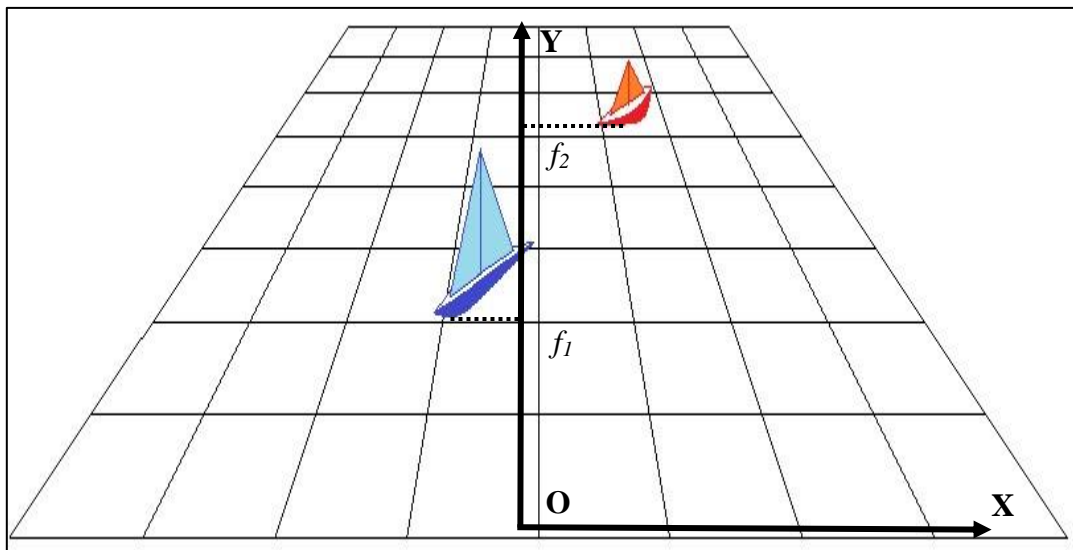
Como se ha explicado previamente, una distancia al *eje-Y* de  $n$  píxeles puede representar distintas distancias a dicho eje dependiendo de la fila en la que nos encontremos. Así pues, para el cálculo de la *coordenada-x* de un píxel tendremos que calcular una función  $F_x$  que depende tanto de la fila  $f$ , como de la columna  $c$ , en la que se encuentra el píxel:

$$x = F_x(f, c) \quad (3.2)$$

En la *figura 3.5(b)* se puede ver claramente cómo el mismo número de píxeles en el *eje-X* (325 píxeles) representan distancias distintas al *eje-Y* según la fila en la que nos encontremos:

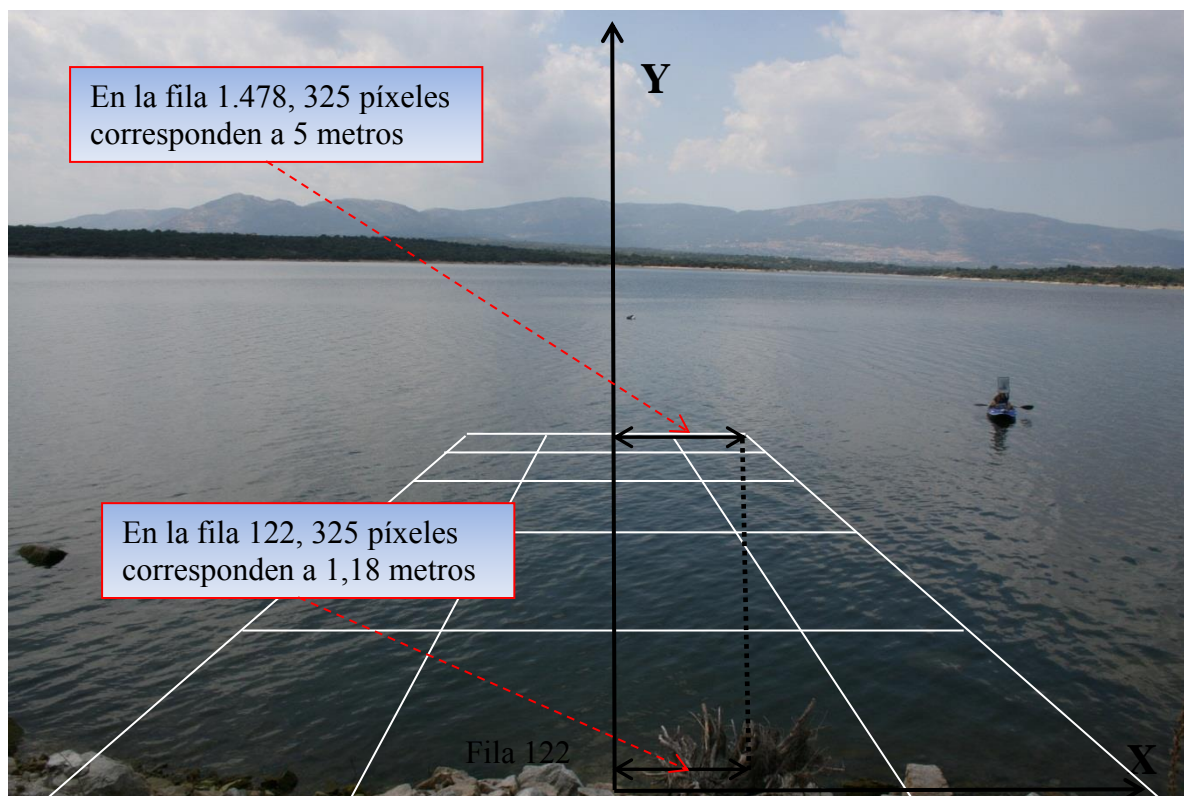


(a)



(b)

**Figura 3.5:** (a) Vista cenital de dos barcos en la superficie del mar. (b) Vista de la misma escena tal cual sería captada por la cámara colocada en el origen de coordenadas. Cuanto más alejado esté el barco de la cámara en la realidad (cuanto mayor sea su coordenada  $-y$ ) más arriba aparecerá en la imagen (mayor será la fila  $f$ ).



**Figura 3.6:** Una misma distancia (325 píxeles) al eje-Y se puede corresponder con una coordenada  $x=5$  o  $x=1,18$  dependiendo de en qué fila nos encontremos.

Al igual que en el caso del cómputo de la *coordenada-y*, utilizamos las mismas imágenes tomadas para su cálculo y procedemos de forma similar. La única diferencia es que en este caso la función que relaciona la *coordenada-x* de la barca depende no sólo de la fila  $f$ , sino también de la columna  $c$ .

## 3.4. Resultados

### 3.4.1. Coordenada-y

Las pruebas se realizaron en el pantano de Valmayor en mayo de 2011. Los equipos utilizados fueron un GPS diferencial Hemisphere GPS A221A220 y una cámara Canon EOS 400D con longitud focal durante la captura de 25 mm.

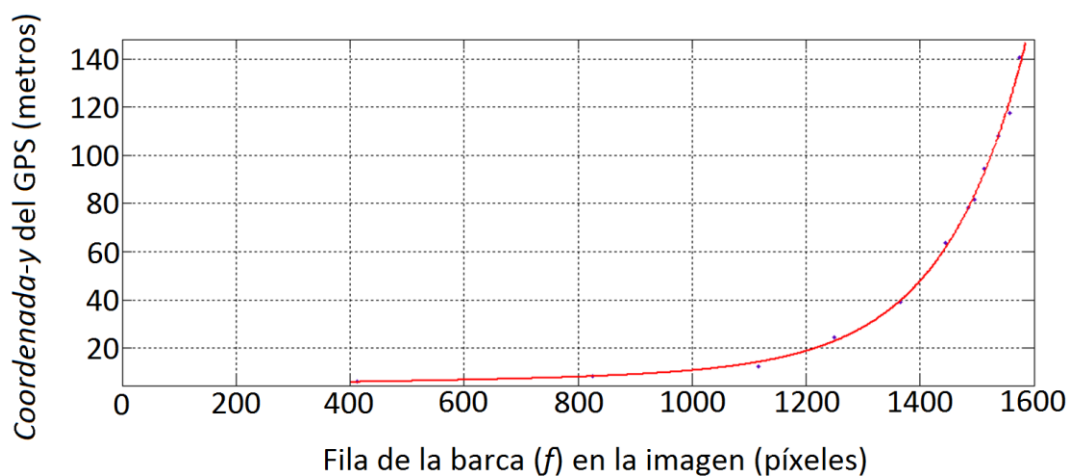
En la tabla 3.1 se muestran las distintas posiciones de la barca en el pantano. Estas posiciones son relativas a las coordenadas GPS en la posición de la cámara. Igualmente se muestran las distintas posiciones de esa misma barca en cada imagen. En dicha tabla aparecen doce imágenes de referencia, si bien se han realizado

aproximadamente cincuenta experimentos con el mismo número de imágenes de referencia, obteniendo resultados similares a los mostrados.

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>Nº imagen</b>	<b>Coordenada-y (metros)</b>	<b>Fila en la imagen, <math>f</math> (píxeles)</b>	<b>Coordenada-x (metros)</b>	<b>Columna en la imagen, <math>c</math> (píxeles)</b>	<b>Ratio píxeles por metro, <math>R=c/x</math> (píxel/m)</b>
1	6,3	413	2,8	742	265,0
2	8,2	825	10	1083	108,3
3	12,3	1.116	-5	-308	61,6
4	24,5	1.249	3,7	146	39,5
5	39,3	1.365	-17,5	-562	32,1
6	63,6	1.445	-9,1	-266	29,2
7	78,3	1.484	39,1	1110	28,4
8	81,8	1.496	49,8	1384	27,8
9	94,5	1.513	47,7	1302	27,3
10	108,0	1.538	-11	-296	26,9
11	117,4	1.558	46,2	1224	26,5
12	140,8	1.574	35,6	935	26,3

**Tabla 3.1:** Imágenes tomadas y datos recogidos con el GPS de la embarcación en las distintas posiciones fotografiadas (columnas B y D), posiciones de la barca en cada imagen (columnas C y E) y relación píxeles por metro (columna F).

Si dibujamos una gráfica de los datos de la *columna B* frente a los datos de la *columna C* obtenemos la curva representada en la *figura 3.7*.



**Figura 3.7:** Relación entre la coordenada-y del objeto y la fila en la que se encuentra en la imagen.

Nuestro objetivo es encontrar la función matemática que relacione la *coordenada-y* y la fila  $f$ , es decir:

$$y = F_y(f) \quad (3.3)$$

Y cuya representación gráfica es la que podemos ver en rojo en la *figura 3.8*. Para ello, hacemos uso de los métodos de cálculo numérico proporcionados por las herramientas *cftools* incluidas en Matlab R2009b (*The Mathworks 2014*).

Dicha herramienta proporciona una serie de curvas tales como polinomios, potencias, series de Fourier y exponenciales entre otras. Dentro de ellas la función que mejor se ajusta a la curva de la *figura 3.8* es la siguiente:

$$y = F_y(f) = 4,603e^{0,000602f} + 0,002255e^{0,006936f} \quad (3.4)$$

Función obtenida con la siguiente fiabilidad de ajuste: SSE: 69.45, R-square: 0.997, Adjusted R-square: 0.9959, RMSE: 2.946.

### 3.4.2. Coordenada-x.

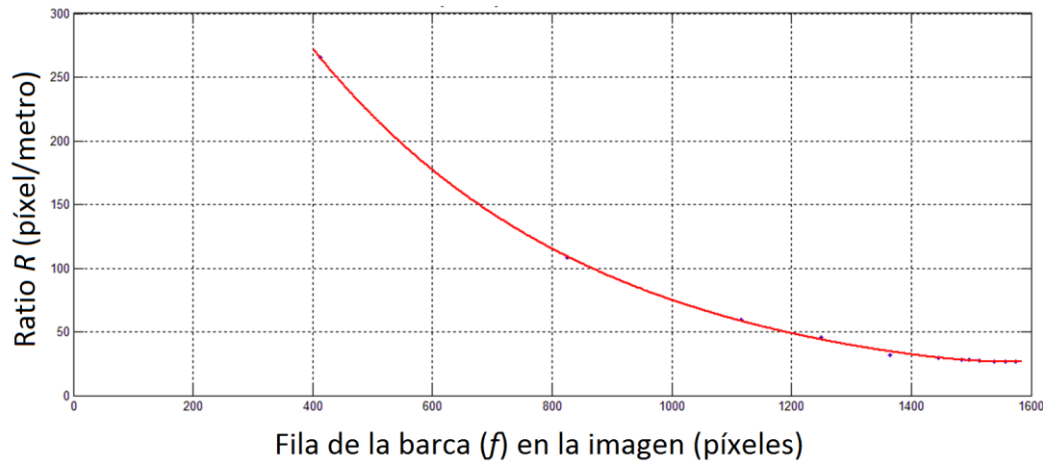
Para calcular  $F_x(f,c)$  a partir de los datos experimentales utilizamos el ratio  $R$  en píxeles por metro:

$$R = c/x \quad (3.5)$$

Es decir, por cada dato obtenido calculamos el ratio  $R$  que nos proporciona el valor de cuántos píxeles representa un metro en esa fila. Ese dato lo añadimos a la tabla de valores (*tabla 3.1, columna F*) y dibujamos la gráfica que nos proporciona la relación del ratio  $R$ , en función de la fila  $f$ , en la que se encuentra (*figura 3.8*).

Al igual que en el apartado anterior, de la gráfica y mediante métodos de aproximación numérica, obtenemos la ecuación que nos relaciona el ratio con la fila, es decir, obtenemos la función  $F_R(f)$  según la ecuación (3.6).

$$R = F_R(f) = 650,4e^{-0,002173f} + 3,01 \times 10^{-8}e^{0,01209f} \quad (3.6)$$



**Figura 3.8:** Relación entre el ratio (píxeles por metro) y la coordenada-x en la que se encuentra el objeto en la imagen.

A partir de las ecuaciones (3.5) y (3.6) obtenemos la *coordenada-x* como función de la fila y la columna como sigue,

$$x = F_x(f, c) = \frac{c}{R} = \frac{c}{F_R(f)} = \frac{c}{650,4e^{-0,002173f} + 3,01 \times 10^{-8}e^{0,01209f}} \quad (3.7)$$

Función obtenida con la siguiente fiabilidad de ajuste: SSE: 34.59, R-square: 0.9993, Adjusted R-square: 0.9991, RMSE: 2.079.

Así pues, una vez calculadas las funciones  $F_y(f)$  y  $F_x(f,c)$  podemos calcular las *coordenadas x-y* de cualquier píxel de la imagen.

La precisión en los resultados depende principalmente de los siguientes factores:

1. Distancia a la que se encuentre el objeto de la cámara. A mayor distancia, menor precisión.
2. Resolución de la imagen. A mayor resolución mayor número de píxeles por metro y por tanto mayor precisión.
3. Precisión en la toma de datos necesaria para la elaboración de la tabla que posteriormente genera las funciones de ajuste  $F_y(f)$  y  $F_x(f,c)$ . Es decir, la precisión en las posiciones GPS del barco y la precisión a la hora determinar la fila y columna en la que se encuentra el barco en la imagen.

4. Sincronización entre la toma de datos GPS y la captura de la imagen en la posición indicada por el GPS.

Una vez obtenidas las funciones  $F_y(f)$  y  $F_x(f,c)$  se procede a cuantificar los errores producidos en cada una de ellas. Para ello, utilizamos otras diez imágenes adicionales cuyas *coordenadas x-y* eran conocidas de antemano gracias al sistema GPS ubicado en la barca. Así pues, disponiendo del “ground truth” procedimos a calcular sus *coordenadas x-y* mediante el uso de las funciones obtenidas. Los resultados se muestran en *las tablas 3.2 y 3.3*.

Número de imagen	Coordenada-y dada por el GPS (m)	Coordenada-y estimada con $F_y(f)$ (m)	Error absoluto (m)	Error relativo (%)
13	6,08	6,16	0,08	1,32
14	7,12	7,06	0,06	0,84
15	7,58	7,44	0,14	1,85
16	9,50	9,70	0,20	2,11
17	17,41	17,83	0,42	2,41
18	22,18	21,61	0,57	2,57
19	67,13	69,48	2,35	3,50
20	82,88	86,28	3,40	4,10
21	105,16	100,01	5,15	4,90
22	133,92	126,96	6,96	5,20
<b>Media</b>			<b>1,93</b>	<b>2,88</b>

**Tabla 3.2:** Errores de la función  $F_y(f)$ , comprobados con las coordenadas reales dadas por el GPS ubicado en la barca.

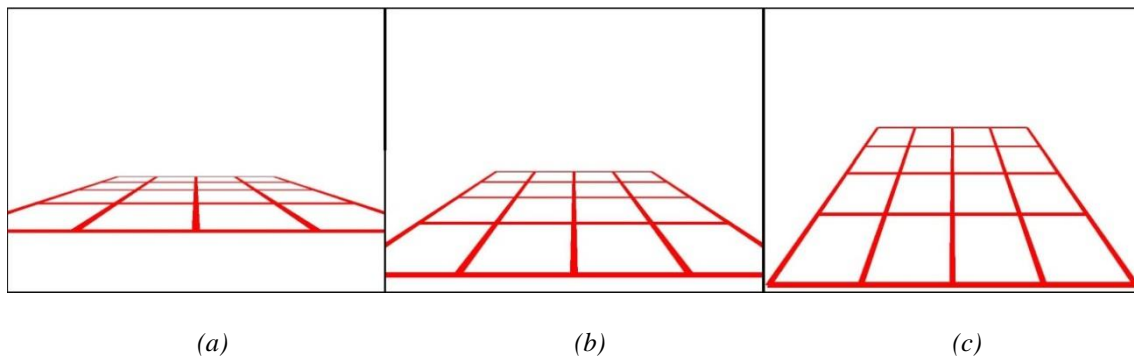
A partir de los datos de las *tablas 3.2 y 3.3* podemos observar cómo los errores aumentan a la vez que aumenta la *coordenada-y*. Esto es debido a que según se incrementa la distancia del objeto en la imagen, el ratio  $R$  decrece, lo que implica que hay menos píxeles de resolución para la misma distancia, como se explicó en la *figura 3.6*.

Por otro lado, el error se puede reducir incrementando la altura de la cámara y modificando su ángulo de inclinación de tal manera que nuestro campo de estudio abarque tanta más área de la imagen como sea posible. Esto es debido a que cuanto más número de filas ocupe nuestro campo de estudio, más filas tendremos para asignar a las distancias y mayor será la precisión, (*figura 3.9*).

Número de imagen	Coordenada-x dada por el GPS (m)	Coordenada-x estimada con $F_x(f,c)$ (m)	Error absoluto (m)	Error relativo (%)
13	-1.34	-1.32	0.01	1.1
14	-1.69	-1.72	0.02	1.4
15	-2.05	-2.09	0.04	1.9
16	-1.55	-1.58	0.03	2.1
17	0.52	0.51	0.01	2.4
18	1.14	1.10	0.03	2.9
19	12.22	12.55	0.33	2.7
20	24.56	25.32	0.76	3.1
21	33.75	32.64	1.11	3.3
22	44.71	43.41	1.30	2.9
		<b>Media</b>	<b>0.37</b>	<b>2.38</b>

*Tabla 3.3: Errores de la función  $F_x(f,c)$ , comprobados con las coordenadas reales dadas por el GPS ubicado en la barca.*

Las figuras 3.8(a),(b) y (c) constituyen un ejemplo pedagógico que muestra lo explicado anteriormente. Se trata de una simulación por computador de una rejilla de 4 x 4 metros. La cámara se ha colocado a 4 metros de distancia de la rejilla. La focal de la lente es de 10 mm., el ángulo de inclinación y la altura de la cámara en cada una de las imágenes son respectivamente:  $0^\circ$  y 0,5 m.,  $6^\circ$  y 1m,  $23^\circ$  y 2m.



*Figura 3.9: Diferentes imágenes simuladas por computador con distintos ángulos de inclinación y altura de la cámara. Según aumenta la altura de la cámara (de figura (a) a figura (c))*

Se han realizado nuevos experimentos con más de 200 imágenes obteniendo en todos los casos precisiones similares a las mostradas en las *tablas* 3.2 y 3.3.

Como se ha mencionado previamente, los resultados del presente algoritmo corresponden a las pruebas realizadas en el pantano de Valmayor (Madrid), sin embargo, también ha sido probado con imágenes de interior, tanto en largos pasillos (40 m) como en habitaciones (3 m). En ambos casos se cumplía la condición necesaria para

poder aplicarlo, es decir, se trataba de superficies planas. Hay que destacar que los errores eran aún menores en estos casos debido a la mayor precisión en la toma de datos así como por la mayor altura de la cámara en relación con el campo de estudio.

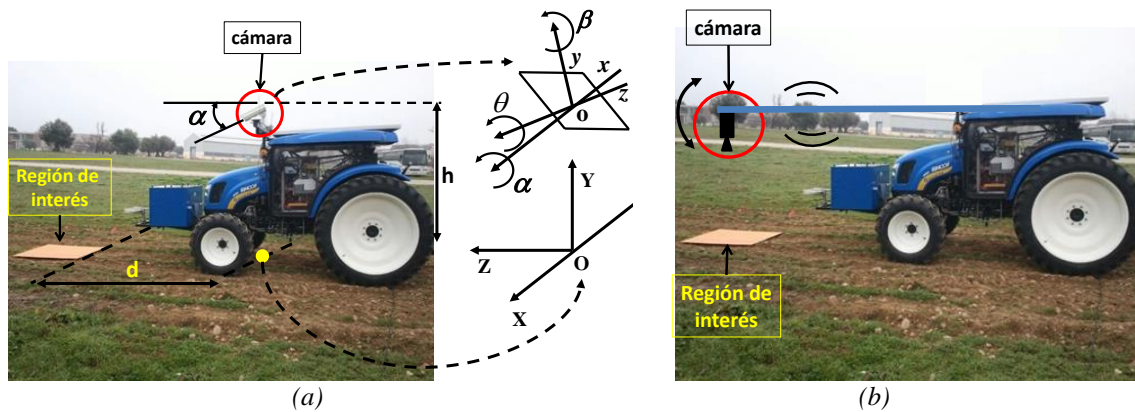


## **AJUSTES DE LA CÁMARA: PRECISIÓN EN LA DETECCIÓN DE LÍNEAS DE CULTIVO Y MALAS HIERBAS**

### **4.1. Introducción**

De todos es conocido el hecho de que un objeto en cualquier imagen se ve mejor y con más resolución cuanto más próximo esté a la cámara en la escena. Sin embargo, se pueden dar situaciones en las que un objeto esté a la misma distancia en dos imágenes distintas y posea más resolución en una que en la otra. Esto es debido a que el mismo objeto puede ser fotografiado variando los ángulos de enfoque de la cámara, sin variar su distancia en cuyo caso estas variaciones afectan directamente a la resolución. Este hecho puede resultar de gran relevancia cuando la resolución de una imagen es un requisito importante para su posterior procesamiento y para los resultados requeridos. En nuestro caso, una vez decidida la ubicación cámara en el tractor, disponemos de cierto margen de libertad en cuanto a la elección de algunos de los parámetros extrínsecos de la propia cámara. Dependiendo de los valores asignados finalmente a dichos parámetros obtendremos una resolución u otra. El presente estudio se ha realizado con el objetivo de determinar qué valores de los parámetros extrínsecos en función de los grados de libertad de la cámara son los que producen imágenes de la zona de interés con la máxima resolución, cuyos resultados se publicaron en *Romeo y col. (2013a)*, *Guerrero y col. (2012c)* y *Montalvo y col. (2011)*. Los grados de libertad a los que nos referimos en este estudio están determinados por los siguientes parámetros geométricos asociados a la cámara, a saber: la altura de la cámara  $h$  con respecto al suelo (ya que disponemos de cierta libertad para desplazarla en su soporte hacia arriba o hacia abajo, situándola a diferentes alturas), el ángulo de inclinación  $\alpha$ , con respecto a una horizontal con referencia el suelo y la distancia a la cámara  $d$ , a la que se encuentra

la zona de cultivo a procesar. En la *figura 4.1* se muestra la geometría y disposición de tales parámetros extrínsecos.



**Figura 4.1:** Disposición de la cámara con una ROI delante del tractor: (a) próxima al centro del centro de masas del tractor con el sistema de coordenadas de referencia. (b) Posición cenital.

La segunda parte de este estudio se centra en dos factores que afectan directamente a la calidad de la imagen. Son tales que, sin un control adecuado, pueden llegar a hacer que la imagen resultante sea completamente inútil para nuestros propósitos. Una vez analizados dichos factores se proponen dos métodos alternativos que solventan los efectos adversos producidos por éstos. Hablamos de dos factores íntimamente relacionados con la iluminación existente en los entornos de exterior, que en nuestro caso concreto afectan de forma crítica a las imágenes. El primero de ellos es la variación constante y aleatoria en la iluminación provocada por el hecho de trabajar en entornos exteriores. Esta variación se compensa, en tiempo real, ajustando el tiempo de exposición en la captura de la imagen durante el cual el CCD de la cámara permanece expuesto a la entrada de radiación lumínica procedente del exterior, formando finalmente la correspondiente imagen. La segunda es el efecto conocido como *vignetting* que consiste en la aparición de un anillo circular alrededor de la imagen, de intensidad decreciente según nos alejamos del centro de la imagen. Este efecto es provocado, entre otros, por el uso de filtros ópticos, identificados normalmente como UV-IR, necesarios para el corte de la radiación ultravioleta e infrarroja presentes ambas en la radiación lumínica procedente del sol. La aparición del anillo concéntrico con la imagen se debe en gran medida a que dichos filtros están fabricados por capas de material filtrante superpuestas, de esta forma, los rayos que inciden en la cámara por el centro lo hacen perpendicularmente y por ello atraviesan el filtro recorriendo su espesor nominal mientras que cuanto más nos alejamos del centro, los rayos atraviesan el filtro con un ángulo creciente y por ello recorren mayores distancias mientras atraviesan el

filtro. El resultado es que al ser “más filtrados” los rayos periféricos habrá mayor corte de rayos UV e IR y por ello aparecerán las zonas exteriores como más oscuras, debido a que pasan con mayor nitidez los rayos cuyas componentes espectrales están relacionadas con las longitudes de onda correspondientes a la banda espectral del verde, esto se explica por el hecho de que estas longitudes de onda se sitúan exactamente entre las correspondientes al UV e IR.

El sistema de visión utilizado en nuestros experimentos consta básicamente de dos partes:

1. El sensor propiamente dicho, constituido en nuestro caso por un dispositivo de tipo CCD.
2. Sistema óptico con su correspondiente longitud focal, en nuestro caso de 10 mm, incluyendo filtros físicos, para controlar la entrada de aquellas longitudes de onda de nuestro interés.

Los parámetros relacionados con las dos partes físicas previas, tales como resolución del sensor CCD, tamaño de los píxeles o la longitud focal, se conocen generalmente como parámetros intrínsecos, en contraposición a los extrínsecos mencionados previamente.

Debido a determinados requerimientos del sistema, algunos de los parámetros intrínsecos y extrínsecos de la cámara son fijos o están previamente definidos. En este apartado nos centramos en los parámetros variables así como en sistemas de compensación de los efectos adversos provocados por una iluminación variable, siempre con el objetivo de obtener la mayor precisión posible en el cálculo del porcentaje de verde y en la detección de las líneas de cultivo.

El presente método ha sido ideado y desarrollado dentro del proyecto RHEA (*RHEA 2014*) perteneciente al VII Programa Marco de la Unión Europea en las áreas tecnológicas nanociencias y nanotecnología, así como en el proyecto AGL (2011) en el área de agricultura. Los resultados iniciales se publicaron con detalle en *Romeo y col. (2013a)*.

## 4.2. Parámetros extrínsecos

El sistema de visión está montado en el tractor y, dependiendo de pequeñas correcciones en su ubicación y orientación, se pueden obtener distintas precisiones en las medidas de densidades de plantas, para una misma distribución de éstas en un área determinada del campo. La *figura 4.1(a)* muestra la configuración de la cámara en el tractor realizada específicamente para el tratamiento de campos de maíz dentro del proyecto RHEA (*RHEA 2014*). En este proyecto un tema de gran interés es el que concierne a las áreas de cultivo que van a ser tratadas debido a que los implementos que aplican el tratamiento abarcan cuatro líneas de cultivo espaciadas 0,75 metros entre sí. El ancho total que debe cubrir la cámara debe ser 3 metros. La solución adoptada para cumplir los requisitos de cobertura en cuanto a ubicación y orientación de la cámara son los mostrados en la *figura 4.1(a)*, sin que por ello sea la única solución posible, ya que siempre se pueden combinar diferentes lentes con distintas longitudes focales, resoluciones y posiciones de la cámara que abarquen igualmente el área requerida con una resolución aceptable para dicho propósito.

Con esta configuración, el origen del sistema de coordenadas referencial *OXYZ* está localizado exactamente en el suelo con los ejes orientados como muestra la *figura 4.1(a)*. A una altura  $h$  del origen  $O$ , se coloca el origen del sistema de coordenadas *xyz* centrado exactamente en el centro del CCD de la cámara. Los ángulos de inclinación ( $\alpha$ ), alabeo ( $\theta$ ) y orientación ( $\beta$ ) definen los tres grados de libertad del plano de la imagen con respecto al sistema de referencia inercial. Como se ha mencionado previamente, son parte de los parámetros extrínsecos que componen la matriz de rotación que define la transformación rotacional entre sistemas de coordenadas espaciales. Un cuarto parámetro extrínseco es  $h$ , que define la altura de la cámara, el cual está incluido en la matriz de traslación también entre sistemas de coordenadas. Aunque la posición de la cámara en el tractor está fijada, aún es posible realizar pequeñas modificaciones en cuanto a su altura  $h$ .

El tractor se mueve a lo largo de las líneas de cultivo y procesa en tiempo real las imágenes, determinando el porcentaje de malas hierbas y localizando las líneas de cultivo por las que avanza. La velocidad del tractor es un parámetro crítico dado que a mayor velocidad del tractor menor coste computacional han de tener los algoritmos de detección de líneas y cuantificación de malas hierbas o en su caso el área de detección

tendrá que enfocarse a mayor distancia del tractor para disponer de más tiempo de procesamiento, con la consiguiente pérdida de precisión.

Idealmente, el mejor ángulo de inclinación sería  $90^\circ$  apuntando hacia abajo (posición cenital). Esta configuración ha sido utilizada para la discriminación entre plantas y malas hierbas en plantaciones de azúcar (*Astrand y Baerveldt 2002*) y su ejecución ha sido descrita en *Baret y col. (2010)*. En nuestro caso, este valor del ángulo de inclinación implicaría que la cámara debería estar literalmente debajo del tractor. Esto conllevaría a su vez que el objetivo no podría alcanzar un ancho de 4 líneas de cultivo pero, aun en ese caso, el procesamiento de las imágenes habría de ser prácticamente instantáneo. Otra solución para mantener un ángulo de inclinación de  $90^\circ$  sería disponer la cámara sobre una barra elevada y ligeramente adelantada sobre el frontal del tractor como muestra la *figura 4.1(b)*. De esta forma, al tener más altura que si estuviese debajo del tractor, el campo de visión sería mayor, siendo factible abarcar las cuatro líneas de cultivo con los objetivos convencionales.

Igualmente, al estar por delante del tractor, se dispondría de un tiempo para el procesamiento de las imágenes, desde que la imagen es tomada hasta que el implemento que ha de aplicar el tratamiento, y que está ubicado en la parte trasera del tractor, pasa por esa misma zona. Sin embargo, por razones mecánicas, esta disposición produciría elevadas oscilaciones de la barra durante la marcha del tractor a través del campo de cultivo. Como resultado de estas oscilaciones se obtendrían imágenes borrosas y difusas que harían prácticamente inútil cualquier procesamiento.

Así pues, una posición aceptable para la cámara, de acuerdo con nuestros requisitos y necesidades, pasa por situarla lo más cerca posible del centro de gravedad del tractor para atenuar al máximo las oscilaciones y vibraciones, con un ángulo de inclinación tal que se cubra el área requerida. La distancia a la que se encuentra la zona a enfocar tiene que permitir el procesado con la mayor precisión posible.

De forma similar, a menor distancia del área de interés ROI (Region Of Interest) respecto del tractor, es decir del origen en el sistema *OXYZ*, mayor precisión en las imágenes. Sin embargo, el área tampoco puede estar demasiado próxima al tractor por dos razones concretas:

1. El tractor posee una caja en el frontal que aloja las baterías de energía, que obligan a elevar ligeramente el ángulo de inclinación para sortear dicho obstáculo en el campo de visión.
2. Como ya hemos mencionado anteriormente, una distancia muy próxima al tractor obligaría a un tiempo de procesamiento de imágenes excesivamente pequeño y por ello inviable.

Respecto a los parámetros extrínsecos dados por los ángulos de alabeo ( $\theta$ ) y orientación ( $\beta$ ) asumimos que están fijados a cero debido a la horizontalidad del terreno (alabeo,  $\theta$ ) y a un correcto guiado del tractor (orientación,  $\beta$ ). Evidentemente, los valores de ambos ángulos no resultan ser exactamente cero debido a la orografía del terreno (alabeo  $\theta$ ) y a los constantes ajustes necesarios en el control de la trayectoria del tractor (orientación  $\beta$ ). Dado que estos últimos factores son ajenos al objetivo del presente trabajo de investigación, nos centramos en los parámetros que afectan directamente a la precisión de las imágenes y de los que sí podemos tener cierto acceso para su control, es decir:

1. El ángulo inclinación de la cámara  $\alpha$ .
2. La altura de la cámara  $h$ .
3. La distancia de la ROI a la base de la cámara  $d$ .

Así pues, este apartado estudia la dependencia de la precisión de las imágenes con respecto al ángulo de inclinación  $\alpha$ , la altura  $h$  y la distancia  $d$ . Estos tres parámetros a su vez están relacionados entre sí, siendo nuestro objetivo encontrar la mejor configuración. Este estudio representa una ampliación del presentado en *Guerrero y col. (2012)*, en el que se proponían como trabajos futuros precisamente los presentados a continuación. En dicho trabajo sólo se analizaba la dependencia de la precisión con el ángulo de inclinación  $\alpha$ , y todos los cálculos se realizaban basándose en escenarios simulados mediante Webots (*Cyberbotics 2014*) al contrario que en este estudio en el que todas las imágenes analizadas son imágenes reales tomadas de campos de cultivo.

Simular escenarios antes de realizar las pruebas en campos reales es una práctica común (*Jones G. y col. 2009a,b*). Dependiendo de la aplicación agronómica específica, algunos sistemas fijan de antemano ciertos parámetros extrínsecos porque las

condiciones y requisitos lo permitan tanto para la detección de cultivos y malas hierbas (*Jiang y col., 2010; Hague y col., 2006 y Bossu, 2009*) como para el seguimiento de líneas (*Astrand y Baerveldt, 2005; Jinlin y Liming, 2010*), incluyendo sistemas de visión estéreo (*Kise y col. 2005*).

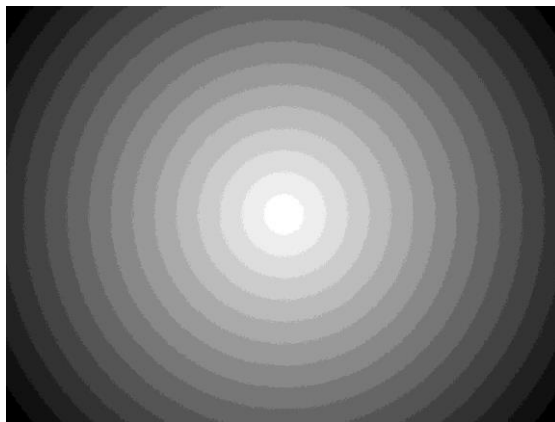
### **4.3. Factores de iluminación que afectan a la calidad de la imagen**

Como se ha mencionado previamente, el procesamiento de imágenes en agricultura es una tarea ciertamente complicada dado que las imágenes son siempre captadas bajo condiciones de iluminación no controladas (*Jeon y col. 2011*). Además, en entornos de exteriores pueden aparecer una gran variedad de condiciones meteorológicas que en definitiva afectan a la calidad de la imagen: elevada iluminación en días soleados, iluminación escasa en días nublados o iluminación variable aleatoria, además de aparición de altos y bajos contrastes, reflejos, cambios bruscos de intensidad solar, etc.

Como añadido, en entornos de exteriores, particularmente en días soleados, la radiación que incide en el sensor le afecta de una forma directa. Esto obliga a la colocación de filtros apropiados que se ajusten a las características de respuesta del sensor. Evidentemente, no existe el filtro perfecto que encaje con las necesidades del sensor filtrando o dejando pasar las longitudes de onda que requiera el sensor en cada momento. Es decir, los materiales fotosensibles de los que están fabricados los sensores y los filtros tienen sus propias curvas de respuesta frente a la radiación solar. Estas curvas son características de cada material y nunca coinciden exactamente las de los filtros con las de los sensores. Es por ello que además se hace necesario la programación en tiempo real que compense esas deficiencias en las operaciones de filtrado de la radiación incidente. La calidad de las imágenes se ve claramente afectada por estos factores, influyendo directamente en los algoritmos de segmentación e indirectamente en la precisión a la hora de identificar líneas de cultivo o plantas en las imágenes, que constituye nuestro tema de interés.

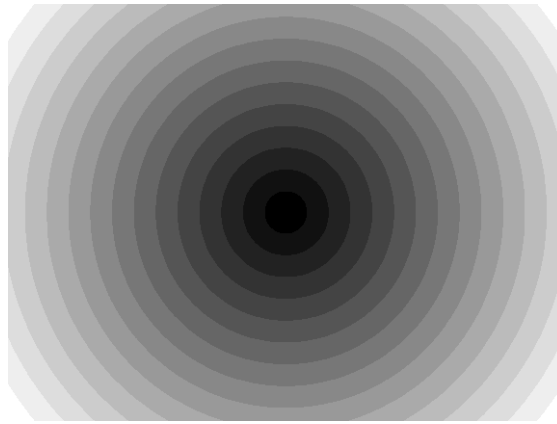
Además, de los parámetros extrínsecos, la precisión de la imagen puede ser estudiada desde el punto de vista de su calidad. En relación a este aspecto, a continuación se describen las siguientes contribuciones:

- *Controlamos la cantidad de luz que impacta el sensor de la cámara.* Debido a que el tamaño de apertura del diafragma se mantiene fijo durante el proceso de captura de la imagen, ya que se trata de un sistema mecánico fijado de antemano, no podemos controlar ni modificar dicho tamaño en tiempo real. Así pues, la cantidad de luz que entra al sensor se controla mediante el tiempo de exposición, cuyo valor sí podemos modificar vía software (*Montalvo y col. 2013b*).
- Corregimos o minimizamos mediante software los efectos causados por el fenómeno conocido con el nombre de *vignetting*. Tal y como se mencionó previamente, consiste en que la luminosidad de la imagen decrece según nos alejamos del centro geométrico de dicha imagen provocando un “anillo visual” en torno a dicho centro. Si fotografiásemos una superficie blanca la imagen obtenida por efecto del *vignetting* sería la que muestra la *figura 4.2*. Evidentemente los efectos del *vignetting* no tienen por qué ser tan exagerados como muestra dicha figura, si bien con este ejemplo se pretende dar una idea de cómo son y qué forma tienen.



**Figura 4.2:** Imagen de una superficie blanca captada con una cámara con un filtro que provoca un *vignetting* muy elevado.

Para corregir el efecto *vignetting* se utiliza un filtro mediante software que contrarresta estos efectos. En la *figura 4.3* se muestra la disposición cualitativa del patrón utilizado para corregir el efecto adverso mediante software. En otras palabras, según nos alejamos del centro de la imagen realzamos la intensidad de los píxeles correspondientes al canal rojo, que son los más afectados por la proximidad al infrarrojo que atenúa el filtro, dicha corrección se realiza de forma proporcional a la distancia al centro de la imagen que se encuentre el píxel.



*Figura 4.3: Patrón aplicado a la imagen mediante software para contrarrestar los efectos del vignetting.*

#### **4.4. Estado del arte**

El uso de sistemas robotizados equipados con sensores basados en visión artificial para tratamientos específicos en agricultura de precisión está en continuo crecimiento. Una práctica común consiste en el procesamiento de imágenes para detección de líneas de cultivo y localización de malas hierbas. Tanto la detección de líneas de cultivo como la localización de malas hierbas son utilizadas para la aplicación de tratamientos selectivos. Además, la detección de las líneas de cultivo es también utilizada para el guiado del tractor. Este tipo de detección es una materia importante en el campo de la visión artificial y el desarrollo de métodos y algoritmos encaminados a ello ha generado numerosos estudios al respecto. Son numerosas las referencias bibliográficas en este sentido, entre las que destacan entre otras muchas las siguientes *López-Granados (2011)*, *Romeo y col. (2012a, 2013a,b)*, *Guerrero y col. (2012a,b, 2013)*, *Gée y col. (2008)*, *Zheng y col. (2009)*, *Montalvo y col. (2012a, 2013a)*, *Guijarro y col. (2011)*, *Burgos-Artizzu y col. (2009)* o *Sainz-Costa y col. (2011)* entre otros.

La determinación del porcentaje de verde frente al de suelo en las imágenes se conoce como “densidad de verde”. El objetivo final es el cálculo de la densidad de verdes en cultivos y malas hierbas, así como las líneas de cultivo a partir de las componentes espectrales de los píxeles en las imágenes. Estas tareas se llevan a cabo mediante una cámara lo más precisa y robusta posible ya que en exteriores, tal y como se ha mencionado previamente, las condiciones de iluminación son altamente variables.

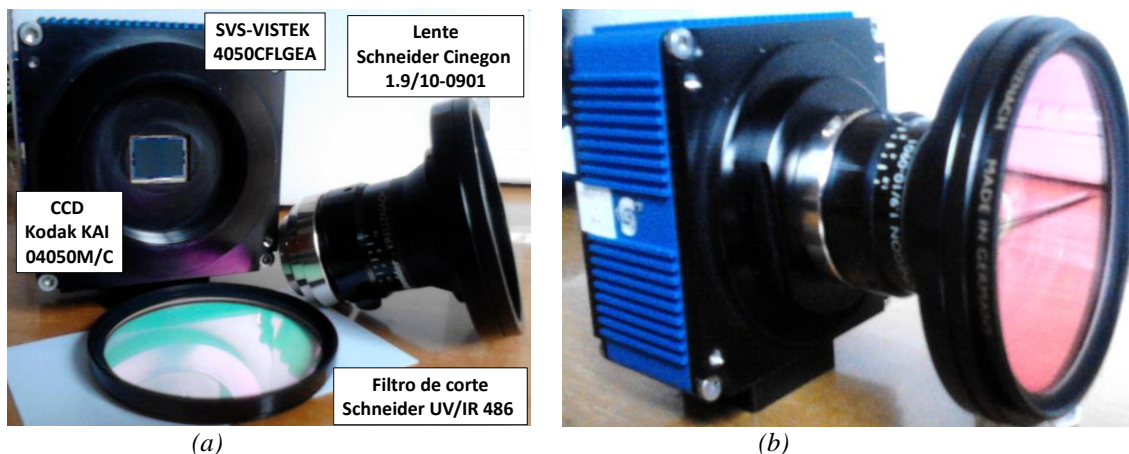
La precisión en estas tareas es crítica para una buena determinación de las densidades y una buena localización de las líneas de cultivo. Está directamente relacionada con los parámetros intrínsecos y extrínsecos de las cámaras antes mencionados (*Fu y col. 1988*), así como algunos factores de iluminación externos no controlables (*Jeon y col. 2011*).

## 4.5. Materiales

La cámara utilizada se compone de tres partes físicas bien diferenciadas que detallamos a continuación:

1. Un sensor CCD con su correspondiente electrónica e interfaces para suministro de energía y conexión al computador.
2. Lentes ópticas.
3. Filtro de corte de luz ultravioleta e infra roja (UV-IR cutting).

En la *figura 4.4(a)*, pueden verse los elementos mencionados por separado y una vez ensamblados en una única pieza *figura 4.4(b)*.



**Figura 4.4:** Sensor CCD, lentes UV/IR y filtro de corte: *(a)* Elementos ópticos por separado. *(b)* Elementos ópticos una vez montados.

El sensor CCD es un sensor Kodak KAI 04050M/C con filtro Bayer para color con rejilla GR, resolución de 2336 x 1752 píxeles y tamaño de píxel de 5,5 x 5,5  $\mu\text{m}$ . Este equipo ha sido diseñado bajo el modelo SVS4050CFLGEA (SVS-VISTEK 2014) el cual es suficientemente robusto para aplicaciones agrícolas.

Esta cámara ofrece numerosas posibilidades para ser controlada externamente en tiempo real:

1. Tiempo de exposición. Para controlar la intensidad de radiación que impacta en el sensor.
2. Control de las ganancias para cada uno de los tres canales: rojo, verde y azul.
3. Establecimiento de la ROI de la imagen.
4. Resolución de la imagen.
5. Información de la temperatura de trabajo de la cámara.

La cámara está conectada a través de una conexión Giga Ethernet a un computador cRIO-9082 dual core a 1,33 GHz con una FPGA, el procesamiento se lleva a cabo con LabView (*National Instruments 2014*). Este sistema de National Instruments es suficientemente robusto en cuanto a características físicas y está específicamente diseñado para ejecuciones en tiempo real, lo cual encaja perfectamente con las aplicaciones agrícolas donde no sólo la velocidad de cómputo es esencial sino también una gran robustez ante las inclemencias atmosféricas y movimientos constantes del tractor. Debido precisamente a los elementos externos (polvo, agua, productos químicos esparcidos) ambos equipos van encapsulados y protegidos de acuerdo con las especificaciones dadas por la normativa IP65. Igualmente, para evitar el sobrecalentamiento de estos equipos en días soleados, debido a las protecciones, van equipados por ventiladores que se activan cuando la temperatura de trabajo se acerca a al límite superior establecido como valor de temperatura recomendable y especificado por el fabricante de la cámara, que en el caso de la cámara antes mencionada es exactamente de +40°C, respecto el límite inferior no se requiere control alguno ya que las labores agrícolas no se realizan con valores de temperatura por debajo de -10°C que es el valor del límite inferior.

El sistema óptico está formado por una lente Schneider Cinegon 1,9/10-0901 (*Cinegon 2013*), con apertura manual de diafragma cuyo rango de valores oscila entre 1,9 y 16 mm. El ajuste del enfoque también es manual y dispone de un bloqueo mecánico, lo cual es muy práctico en este tipo de aplicaciones ya que las propias vibraciones del tractor así como las oscilaciones provocadas por la orografía del terreno, cuando el tractor está en movimiento, acabarían por desenfocar la imagen de forma aleatoria. Este objetivo es válido para sensores de hasta una pulgada (de diagonal) en el

sensor CCD, es decir, un círculo máximo de imagen de 16 mm de diámetro. Está equipado con una montura-F que puede ser adaptada a una montura-C. La focal de la lente es fija, con valor de 10 mm. El campo de visión se sitúa por encima de los 50° y la distancia nominal para captura de objetos varía desde el infinito hasta los 7,5 mm, lo cual nos permite abarcar los 3 metros de anchura necesarios para cubrir las cuatro líneas de cultivo previstas. Los rangos espectrales de respuesta del sensor CCD varían desde los 400 nm hasta los 1000 nm, es decir, el espectro visible y el infrarrojo cercano NIR (Near Infra Red). El sistema óptico descrito proporciona unas imágenes de trabajo con proyección de perspectiva (al igual que las cámaras comerciales convencionales).

Como ya hemos mencionado, nuestro sistema trabaja bajo condiciones atmosféricas adversas y cambiantes donde la iluminación contiene una alta componente de radiación infra roja. El sensor es altamente sensible a la radiación NIR y algo menos a la radiación ultravioleta (UV). En la práctica la radiación NIR contamina en exceso los tres canales espectrales (rojo, verde y azul), si bien resulta mucho más acusada en el rojo, produciendo imágenes en tonos rojizos (véase *figura 4.5(a)*). Esto hace que la identificación de malas hierbas y líneas de cultivo sea prácticamente imposible ya que no es posible segmentar apropiadamente la imagen extrayendo aquellos píxeles con un determinado porcentaje de verdes. Para evitar este efecto no deseado, el sistema va equipado con un filtro de corte UV/IR Schneider (*Schneider 2013*). De acuerdo con su curva de operación las longitudes de onda por debajo de los 370 nm y por encima de los 760 nm son bloqueadas, es decir, bloquea tanto la radiación NIR como la UV. No obstante, a pesar del bloqueo de esta radiación, o quizá debiéramos decir, como efecto secundario del corte de esta radiación, aparece el efecto *vignetting*, el cual requiere una corrección que detallamos más adelante. En efecto, cuando se bloquea en la entrada parte del infrarrojo, el canal más afectado es precisamente el rojo, debido a que se llega a bloquear parte de la entrada del canal rojo, como consecuencia de esto aparecen tonalidades con predominio del verde (también del azul), esto es lo que ocurre en la imagen mostrada en la *figura 4.5(b)*, cuyo efecto se pretende corregir.

Se han capturado más de 2000 imágenes entre Abril/Mayo/Junio y Julio de los años de 2011, a 2013 en campos experimentales de la Poveda pertenecientes al Instituto de Agricultura Sostenible del Consejo Superior de Investigaciones Científicas en Arganda del Rey (Madrid), en los campos experimentales de la Escuela Técnica

Superior de Ingenieros Agrónomos de Madrid. Las últimas imágenes tomadas en noviembre de 2012 se obtuvieron adaptando una explanada repleta de hierba natural mediante un proceso de arado con ayuda manual a distancias simuladas para conseguir valores similares a las de los cultivos de maíz, es decir a 75cm. El motivo es debido a que en esta época del año no existen cultivos reales de maíz.



**Figura 4.5:** (a) Imagen de líneas de cultivo tomada sin filtro de corte UV/IR; (b) la misma imagen con filtro de bloqueo UV/IR.

Para calcular la precisión de las imágenes tomadas, se colocó una cartulina de un metro de ancho por un metro de largo en la zona de nuestro interés a modo de ROI. La cartulina era de color naranja intenso para su mejor segmentación a la hora de contar los píxeles naranjas presentes en la imagen.

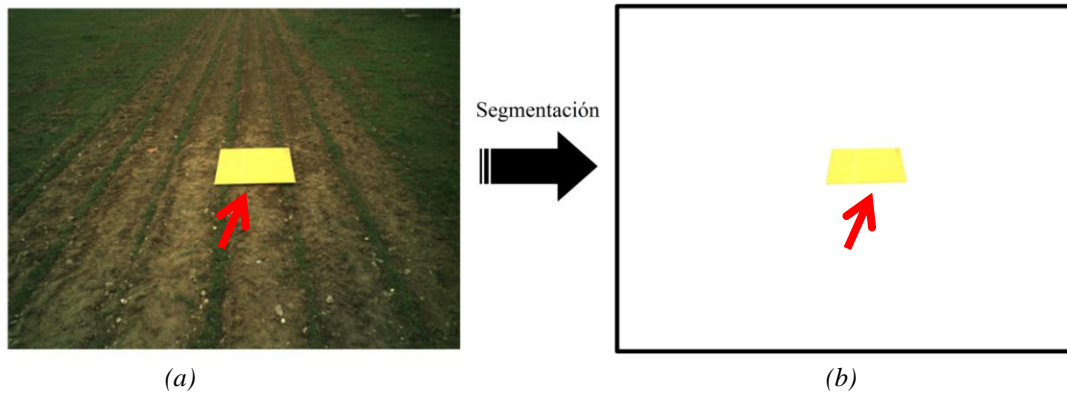
## 4.6. Descripción del método

### 4.6.1. Parámetros extrínsecos

Para cuantificar la resolución de la ROI obtenida en cada una de las imágenes capturadas, se procede a una segmentación del color naranja intenso, generado por la cartulina dispuesta a tal efecto, dicho color no está presente en los campos de cultivo, de forma que una vez eliminados el resto de píxeles de la imagen excepto los pertenecientes a la cartulina, se realiza el cómputo del número de píxeles mediante la correspondiente aplicación desarrollada en Matlab R2009b (*The Mathworks 2014*).

La segmentación es relativamente sencilla dado que el color naranja intenso es fácilmente discernible respecto al resto en campos de cultivo por la razón expuesta previamente. El resultado de dicha segmentación puede verse en la *figura 4.6*. A mayor número de píxeles naranjas  $n$ , presentes en la imagen segmentada mayor resolución en la ROI y por ello mayor precisión en cualquier procesado posterior que realicemos. Este

es el criterio utilizado para cuantificar la precisión en función de los parámetros extrínsecos estudiados.



**Figura 4.6:** (a) Región de interés del estudio, ROI (definida por la cartulina naranja). (b) Número de píxeles ocupados por la ROI en la imagen una vez eliminados los píxeles “no-naranjas” en (a).

Para el estudio de la dependencia de la precisión de las imágenes con los parámetros extrínsecos  $\alpha$ ,  $h$  y  $d$  se han tomado 125 imágenes ( $= 5 \times 5 \times 5$ ), es decir, se han seleccionado imágenes con 5 ángulos de inclinación diferentes  $\alpha$ , 5 alturas de cámara diferentes  $h$  y 5 distancias de ROI a la base de la cámara diferentes  $d$ , combinadas entre sí.

El procedimiento seguido para la captura de las 125 imágenes, variando los tres parámetros es el siguiente:

for  $h = h_1$  to  $h_5$  do

for  $d = d_1$  to  $d_5$  do

for  $\alpha = \alpha_1$  to  $\alpha_5$  do

- Captura de la imagen  $I_{hda}$
- Identifica la cartulina y aplica la segmentación basada en las componentes espectrales RGB.
- Calcula la resolución de la imagen contando el número de píxeles  $n$ , pertenecientes a la cartulina.

end;

end;

end;

Con este procedimiento se crea una tabla con los valores de  $n$  obtenidos para cada  $I_{hda}$ , a continuación se analizan los resultados y se establece la relación de la precisión con dichos valores, cuya representación se muestra en la *sección 4.7 “Resultados”*.

## 4.6.2. Factores de iluminación

Como se ha comentado previamente, tenemos dos factores principales no controlados que afectan a la calidad de la imagen y consecuentemente a la precisión de los resultados una vez procesada, a saber: tiempo de exposición y efecto *vignetting*. A continuación describimos los procedimientos seguidos en cada caso.

### 4.6.2.1. Tiempo de exposición

Colocamos un panel con cuatro colores enfrente del tractor de tal forma que dicho panel aparezca en la parte inferior de la imagen. Llegados a este punto damos por optimizados los parámetros extrínsecos en cuanto a la precisión de las imágenes se refiere y nos centramos en la calidad de las mismas.

El procedimiento consta de los siguientes pasos:

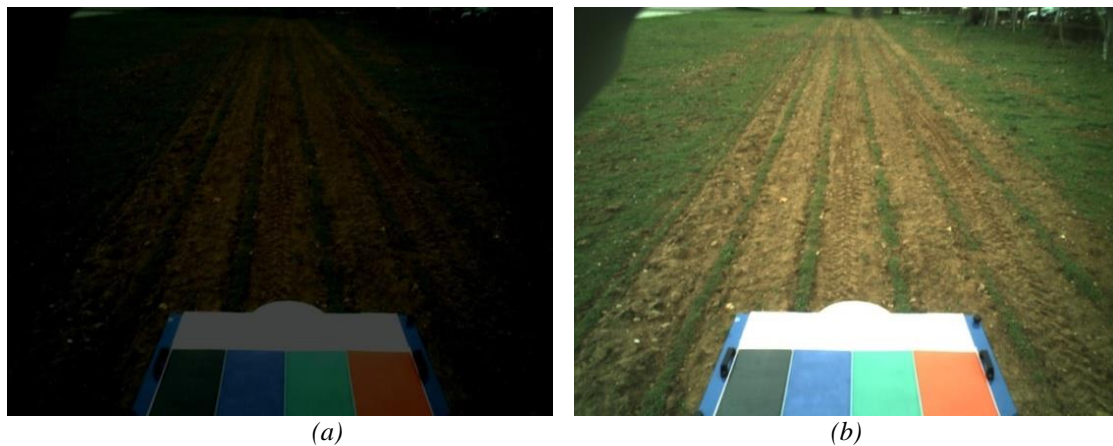
1. Captura de la imagen.
2. Muestreo de píxeles pertenecientes a la zona blanca del panel, calculando los valores medios de las tres componentes espectrales RGB, es decir, los valores medios del rojo, verde y azul.
3. Ajuste del tiempo de exposición  $E_t$ , de tal forma que el mayor de los tres valores promedio  $H = \max\{R, G, B\}$  quede dentro de un intervalo para asegurar que el sensor CCD está suficientemente excitado sin llegar a saturarse. En nuestro sensor con 8 bits de resolución por canal, el valor máximo es  $M = 2^8 = 255$ ; así pues, el intervalo dentro del cual ha de estar  $H$  es el siguiente:  $aM \leq H < bM$ , donde  $a$  y  $b$  definen los límites inferior y superior. En nuestros experimentos queda verificado que con los siguientes valores  $a = 0,90$  y  $b = 0,98$  es suficiente.
4.  $E_t$  representa el tiempo de exposición anterior y se ajusta de la siguiente manera:
  - si  $H < 0,90M$  entonces  $E_t = (1+p)E_{t-1}$
  - si  $H > 0,98M$  entonces  $E_t = (1-p)E_{t-1}$
  - en caso contrario  $E_t$  no necesita actualizarse

Siendo  $0 < p < 1$  representa el salto entre sucesivos ajustes, que en nuestro caso se ha fijado en 0,2.

5. Si después del ajuste  $H$  no está dentro del intervalo especificado, se toma una nueva imagen con el último  $E_t$  actualizado y se repiten los pasos anteriores 1 a 4.

La *figura 4.7(a)* muestra una imagen captada con un tiempo de exposición  $E_t$  de 20 ms, donde fácilmente puede apreciarse la necesidad de ajustar la intensidad de colores.

En la *figura 4.7(b)* la misma imagen es capturada con un tiempo de exposición  $E_t$  de 36 ms, observándose claramente la diferente tonalidad entre ambas.



**Figura 4.7:** Imágenes obtenidas de una misma escena utilizando dos tiempos de exposición diferentes. (a)  $E_t = 20$  ms; (b)  $E_t = 32$  ms.

#### 4.6.2.2. Corrección del *vignetting*

Según las especificaciones del fabricante Schneider, el filtro de corte UV/IR 486 (*Schneider 2013*) está basado en lo que se conoce con el nombre de tecnología de capas finas. Dicho filtro presenta más de treinta capas en uno de sus lados y una capa multi-resistente en el otro. El ángulo de incidencia de los rayos en la periferia del filtro es mayor que en el centro y deben recorrer mayores distancias a lo largo de las diferentes capas del filtro. Este efecto es tanto más acusado cuanto menor sea la longitud focal, es decir, lentes con amplios ángulos de vista. En nuestro caso, con una lente de 10 mm de focal es suficiente para padecer los efectos del *vignetting*.

El efecto es que el color real es desplazado hacia el color verde y azul a costa del color rojo. Para corregir este efecto se ha diseñado la imagen patrón  $P$ , mostrada en la *figura 4.8*, que coincide con el mostrado previamente en la *figura 4.3*, si bien ahora remarcado en rojo para simular el efecto sobre dicho canal. Los valores de  $P$  varían entre  $[0, 1]$ . En el centro de  $P$  ( $c_x, c_y$ ) el valor es cero y adquiere su valor máximo de la unidad en las cuatro esquinas. El tamaño de  $P$  coincide con el de la imagen y los valores

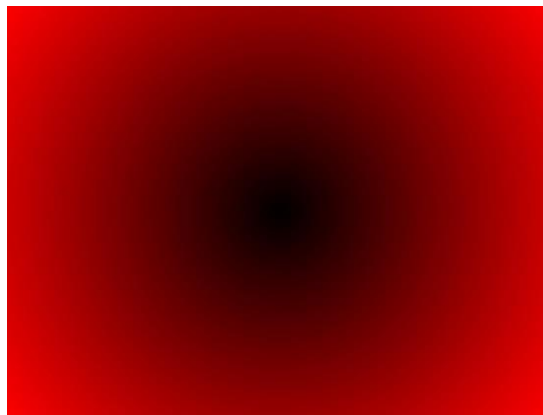
de cada píxel  $i$  ubicado en las *coordenadas*  $x$ - $y$  se calculan según las ecuaciones (4.1) y (4.2).

$$d(x, y) = \sqrt{(x - c_x)^2 + (y - c_y)^2} \quad (4.1)$$

Así pues, para obtener los valores corregidos  $R_c$ ,  $G_c$  y  $B_c$  de los tres canales  $R$ ,  $G$  y  $B$  se procede como sigue:

$$R_c = (1 + K_r P) \times R; \quad G_c = (1 + K_g P) \times G; \quad B_c = (1 + K_b P) \times B \quad (4.2)$$

Donde “ $\times$ ” denota multiplicación píxel a píxel en lugar de multiplicación de matrices;  $K_r$ ,  $K_g$  y  $K_b$  representan los valores de compromiso entre correcciones basándonos en el comportamiento del filtro de corte Schneider UV/IR 486 (Schneider 2013). Hemos podido comprobar con alto nivel de satisfacción que los siguientes valores son apropiados en nuestros experimentos:  $K_r = 0,3$ ;  $K_g = 0$  y  $K_b = 0$ . Obtenidos mediante pruebas de ensayo y error tras diversos experimentos. Es decir, sólo el canal rojo es corregido en nuestra aplicación de extracción de líneas de cultivo.



**Figura 4.8:** Imagen del filtro por software utilizado para la corrección del vignetting. Obsérvese que en el centro geométrico de la imagen no se aplica corrección alguna (valor 0) y en las esquinas se aplica la máxima corrección (valor 255).

## 4.7. Resultados

El presente estudio presenta los resultados obtenidos mediante el método descrito en la *sección 4.6 “Descripción del método”*. Concretamente son los relativos a la precisión de las imágenes para la detección de líneas de cultivo y malas hierbas dentro del proyecto RHEA (RHEA 2014) en campos de maíz. Está centrado en dos grupos de factores que afectan a dicha precisión. Tal y como se describió en dicha

sección, por un lado estudia la influencia de tres parámetros extrínsecos críticos, el ángulo de inclinación  $\alpha$ , la altura de la cámara  $h$ , y la distancia de la ROI a la base de cámara  $d$ , en relación a la resolución de las imágenes y por otro en el control de efectos adversos provocados por la iluminación que afectan a la calidad de la imagen como son el tiempo de exposición y un efecto secundario provocado por los filtros solares denominado *vignetting*.

#### 4.7.1. Análisis de los parámetros extrínsecos

La *tabla 4.1* muestra el número de píxeles  $n$ , obtenidos para la ROI después de aplicar el método descrito en la *sección 4.6.1* “Parámetros extrínsecos” en un conjunto de 125 imágenes. Cada imagen  $I_{\alpha h d}$  es captada con cinco valores distintos del ángulo de inclinación  $\alpha$ , otros cinco de la altura de la cámara  $h$  y otros cinco diferentes de la distancia a la ROI  $d$ .

Para mayor facilidad de comprensión los resultados de la *tabla 4.1* se muestran gráficamente en la *figura 4.9*, con los 125 valores obtenidos para  $n$ . Como podemos ver en dicho gráfico, existen cinco grupos de cinco valores cada uno (marcados con un círculo rojo) que corresponden a los máximos valores de  $n$ . Todos estos valores fueron obtenidos a una distancia  $d = 3$  metros lo que nos lleva a la primera conclusión de nuestro estudio (conclusión bastante intuitiva pero no obvia), es decir: “*Cuanto más próxima esté la ROI a la cámara mayor es la resolución*”.

Hay que tener en cuenta que hemos llegado a esta conclusión analizando unos rangos de valores concretos para  $\alpha$ ,  $h$  y  $d$ . Sin embargo, como veremos más adelante, se pueden dar situaciones en las que sea más conveniente, en términos de resolución, alejar la ROI de la base de la cámara dependiendo de  $\alpha$  y  $h$ , *figura 4.13*. Es por ello que, aunque la conclusión anterior parezca ser bastante intuitiva, no posee un carácter de aplicación general sino que solamente aplica para los rangos de valores de los parámetros estudiados.

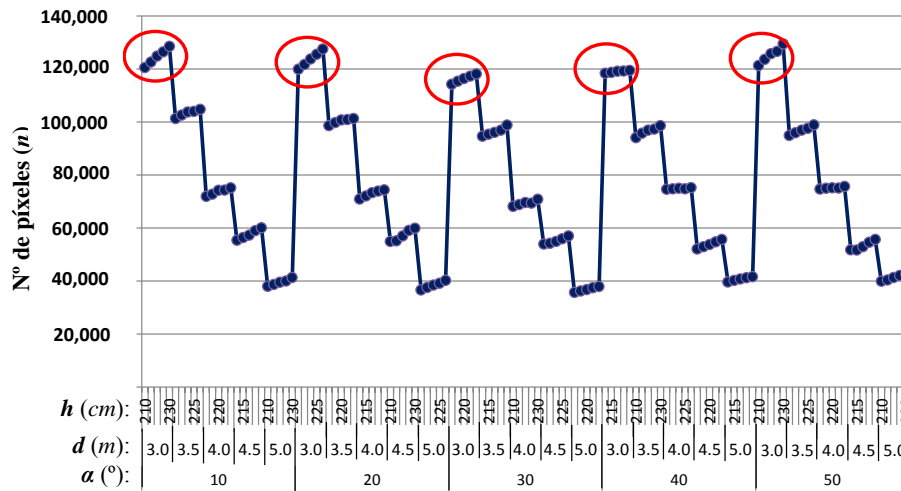
Una vez conocido que la distancia  $d$  es un parámetro crítico para la precisión, fijamos la distancia de la ROI a 3 metros y vemos cómo varía  $n$  con los otros dos parámetros. De esta forma, con la distancia fijada en  $d = 3\text{m}$  dibujamos la variación de  $n$  según los distintos valores de  $\alpha$  y  $h$ , *figura 4.10*.

$\alpha = 10^\circ$			$\alpha = 20^\circ$			$\alpha = 30^\circ$			$\alpha = 40^\circ$			$\alpha = 50^\circ$		
$d$ (m)	$h$ (cm)	$n$	$d$ (m)	$h$ (cm)	$n$	$d$ (m)	$h$ (cm)	$n$	$d$ (m)	$h$ (cm)	$n$	$d$ (m)	$h$ (cm)	$n$
3	210	120.456	3	210	119.872	3	210	114.152	3	210	118.338	3	210	121.236
	215	122.547		215	121.622		215	115.351		215	118.645		215	123.568
	220	124.754		220	123.682		220	116.321		220	119.078		220	125.698
	225	126.354		225	125.462		225	117.244		225	119.176		225	127.425
	230	128.452		230	127.442		230	118.089		230	119.357		230	129.423
3,5	210	101.239	3,5	210	98.589	3,5	210	94.523	3,5	210	94.021	3,5	210	94.865
	215	102.569		215	99.745		215	95.325		215	95.685		215	95.862
	220	103.698		220	100.695		220	95.986		220	96.852		220	96.899
	225	103.967		225	100.885		225	96.854		225	97.246		225	97.585
	230	104.693		230	101.210		230	98.752		230	98.563		230	98.865
4	210	71.895	4	210	70.912	4	210	68.108	4	210	74.594	4	210	74.625
	215	72.698		215	71.992		215	68.812		215	74.729		215	74.987
	220	74.147		220	73.245		220	69.543		220	74.947		220	75.125
	225	74.236		225	73.856		225	69.332		225	74.752		225	75.012
	230	75.100		230	74.342		230	70.811		230	75.219		230	75.652
4,5	210	55.326	4,5	210	54.856	4,5	210	53.865	4,5	210	52.023	4,5	210	51.745
	215	56.314		215	55.123		215	54.169		215	52.869		215	51.658
	220	57.259		220	56.985		220	54.896		220	53.695		220	52.896
	225	58.965		225	59.001		225	55.896		225	54.754		225	54.585
	230	59.996		230	59.896		230	56.987		230	55.625		230	55.632
5	210	37.989	5	210	36.552	5	210	35.648	5	210	39.602	5	210	39.874
	215	38.654		215	37.472		215	36.182		215	40.226		215	40.325
	220	39.541		220	38.442		220	36.762		220	40.713		220	41.256
	225	39.987		225	39.152		225	37.421		225	41.160		225	42.015
	230	41.253		230	40.232		230	37.895		230	41.581		230	42.569

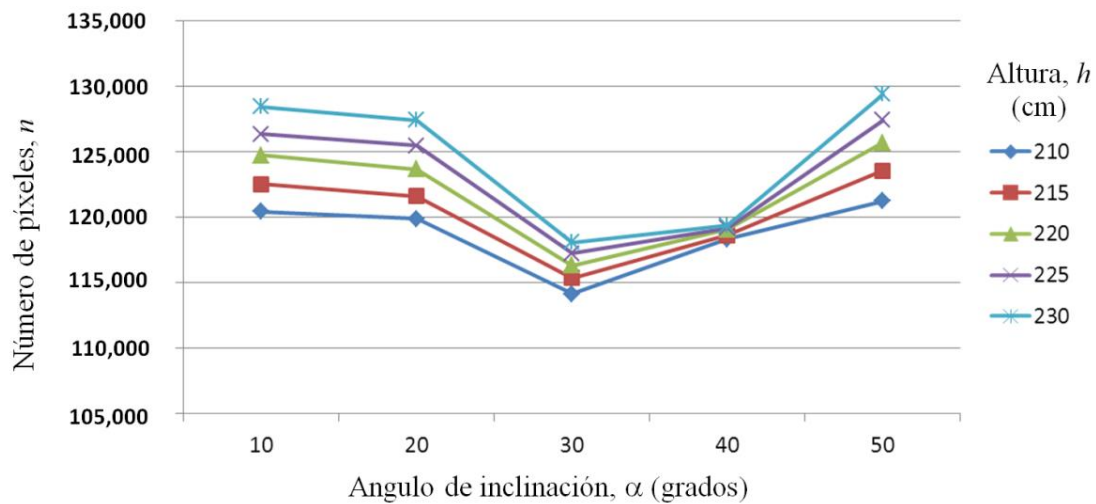
**Tabla 4.1:** Número de píxeles  $n$ , en la ROI obtenidos con la combinación de diferentes ángulos de inclinación  $\alpha$ , distintas alturas de la cámara  $h$ , y distintas distancias de la ROI a la cámara  $d$ .

A partir de la *figura 4.10* podemos observar que los mejores ángulos de inclinación, en términos de precisión, son aquellos en los que la ROI aparece completa en la parte superior de la imagen ( $\alpha = 50^\circ$ ), *figura 4.11(a)*, o bien en la parte inferior de la imagen ( $\alpha = 10^\circ$ ), *figura 4.11(b)*, mientras que el peor ángulo de inclinación en términos de precisión se corresponde con aquel en el que la ROI aparece en el centro de la imagen ( $\alpha = 30^\circ$ ), *figura 4.11(c)*. Las imágenes de la *figura 4.11* fueron captadas con una altura de cámara  $h = 215$  cm y la ROI a una distancia  $d = 3$  m. La conclusión que acabamos de obtener ya fue esbozada en *Guerrero y col. (2012b)*, aunque en esa ocasión sólo se consideraba el ángulo de inclinación  $\alpha$ , en exclusiva sin la intervención

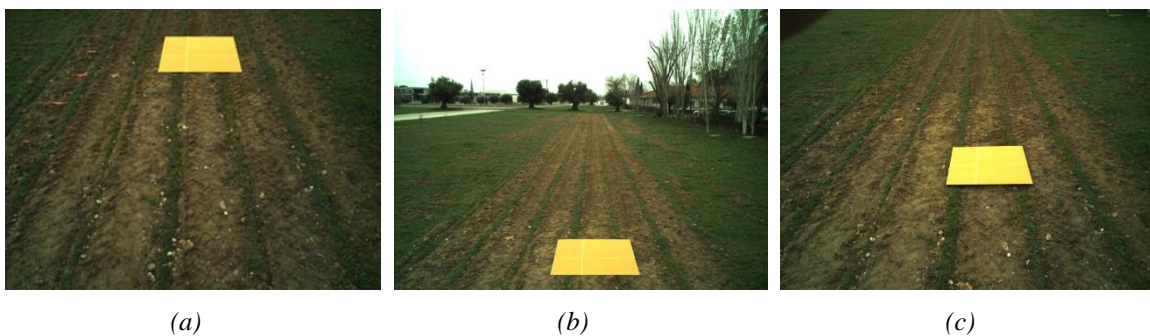
de la altura  $h$ . Así pues, después de un estudio más exhaustivo llegamos a la misma conclusión con respecto al ángulo inclinación  $\alpha$ .



**Figura 4.9:** Variación del número de píxeles  $n$ , en función del ángulo de inclinación  $\alpha$ , de la altura de la cámara  $h$ , y de la distancia de la ROI a la base de la cámara  $d$ .



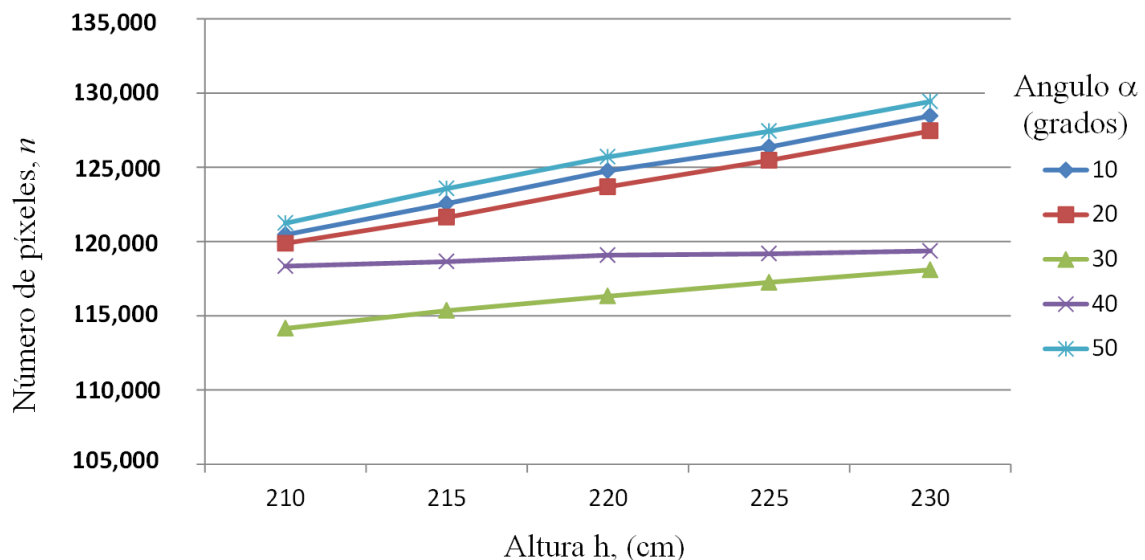
**Figura 4.10:** Número de píxeles  $n$ , en función del ángulo de inclinación  $\alpha$ , para cinco alturas diferentes  $h$ , a una distancia fija  $d=3$  m.



**Figura 4.11:** Distintas imágenes de la ROI (la cartulina naranja) obtenidas con una altura  $h=215$  cm y a una distancia  $d=3$  m bajo tres ángulos distintos de inclinación: (a)  $\alpha=50^\circ$ ; (b)  $\alpha=10^\circ$  y (c)  $\alpha=30^\circ$ .

Con respecto a la altura  $h$ , en la *figura 4.12* podemos ver que el mayor número de píxeles en la ROI  $n$ , se obtiene para una altura de  $h = 230$  cm. Este resultado parece que contradice la intuición ya que aparentemente a menor altura, nos encontraríamos más próximos a la ROI y por ello la resolución y la precisión deberían ser mayores de acuerdo con la conclusión que acabamos de extraer unos párrafos atrás al observar la *figura 4.9*.

Para explicar estos resultados mostramos nuevamente de forma gráfica el número de píxeles  $n$  frente a  $\alpha$  y  $h$ , solo que esta vez colocamos en el *eje-X* de la gráfica la altura  $h$ , *figura 4.12*. Claramente podemos ver cómo la precisión mejora según aumentamos la altura de la cámara en el tractor. Esto es debido a las propiedades geométricas inherentes a la disposición de la cámara. La *figura 4.13* muestra un ejemplo pedagógico para clarificar este hecho. Si colocamos la cámara a muy poca altura, esto es, muy próxima al suelo, obtendríamos la imagen de la *figura 4.13(a)*, sin embargo, manteniendo el mismo ángulo de inclinación  $\alpha$ , y la misma distancia de la ROI,  $d$ , para una altura mayor, la imagen obtenida es la que aparece en la *figura 4.13(b)*. Como puede observarse, el área ocupada por la ROI en la *figura 4.13(b)* es considerablemente mayor que el área ocupada en la *figura 4.13(a)*, lo que se traduce en un mayor número de píxeles  $n$ , y por tanto una mayor precisión.



**Figura 4.12:** Número de píxeles  $n$ , según la altura  $h$ , para cinco ángulos de inclinación diferentes  $\alpha$ , a una distancia fija  $d = 3$  m.



**Figura 4.13:** Imágenes de la misma ROI a la misma distancia y con el mismo ángulo de inclinación. (a) Con la cámara prácticamente a nivel del suelo; (b) Con la cámara a una distancia mayor del suelo que en la imagen (a). Obsérvese cómo la precisión es mayor en el caso (b), es decir, a una mayor altura, a pesar de estar a mayor distancia de la ROI.

Para darnos cuenta de la importancia de elegir los parámetros correctos a la hora de obtener la máxima precisión en un área de la imagen determinada veamos el número máximo de píxeles  $n_{max}$ , y el mínimo  $n_{min}$ , obtenido en nuestro estudio:

$$n_{max} = 129.423 \text{ píxeles (con } \alpha = 50^\circ, h = 230 \text{ cm, } d = 3\text{m).}$$

$$n_{min} = 35.648 \text{ píxeles (con } \alpha = 30^\circ, h = 210 \text{ cm, } d = 5\text{m).}$$

Lo que significa que para la misma cámara, en las mismas condiciones de trabajo, una misma ROI puede ser representada con un número bien distinto de píxeles dependiendo de los valores de los parámetros  $\alpha$ ,  $h$  y  $d$  elegidos. Esto puede ser crítico en determinadas aplicaciones agrícolas ya que si un tipo de ajustes paramétricos proporciona un área de la ROI con  $n = 129.423$  píxeles mientras que con otro ajuste en las mismas condiciones se obtiene un valor de  $n = 35.648$  píxeles estaríamos hablando de una pérdida de resolución del 72,5%.

Hemos estudiado la mejor configuración teórica de los parámetros extrínsecos llegando a un compromiso entre ellos. Ahora bien, en nuestro caso, considerando que el tractor está en movimiento a una velocidad de 4 km/h y que nuestra ROI mide 5 metros de profundidad por 4,5 metros de ancho una buena opción para los parámetros resulta ser:  $\alpha = 20^\circ$ ,  $h = 220 \text{ cm}$  y  $d = 4 \text{ m}$ , ya que estos valores nos permiten obtener una ROI sobre el terreno lo suficientemente alejada como para disponer del tiempo necesario para procesar la imagen, pero lo suficientemente cercana como para disponer de una resolución aceptable de la misma en la imagen. Esta selección de valores, según nuestro estudio, nos proporcionan una resolución de  $n = 56.985$ . Si comparamos esta resolución

con la máxima posible que podríamos obtener en esas circunstancias, esto es, con un valor  $n = 129.423$ , nos damos cuenta que debido a restricciones ajenas a la cámara, provocadas por la necesidad del procesamiento en tiempo real, estamos renunciando a un 56% de precisión. Sin embargo, si no hubiésemos tenido en cuenta la importancia de estos parámetros, una peor elección podría habernos hecho perder hasta un 72,5% de resolución. Así pues, la precisión con distintas disposiciones que nos permitirían trabajar en tiempo real puede llegar a oscilar un 16,5% dependiendo de la elección de parámetros realizada.

Aparte de los parámetros considerados en este estudio existen otros también de naturaleza extrínseca a tener en cuenta con una influencia que puede llegar a ser importante dependiendo de la aplicación. Nos referimos a los ángulos de orientación y alabeo. Estos ángulos han de ser considerados si las circunstancias lo requieren, por ejemplo cuando el tractor trabaja en terrenos con pendiente ladeada.

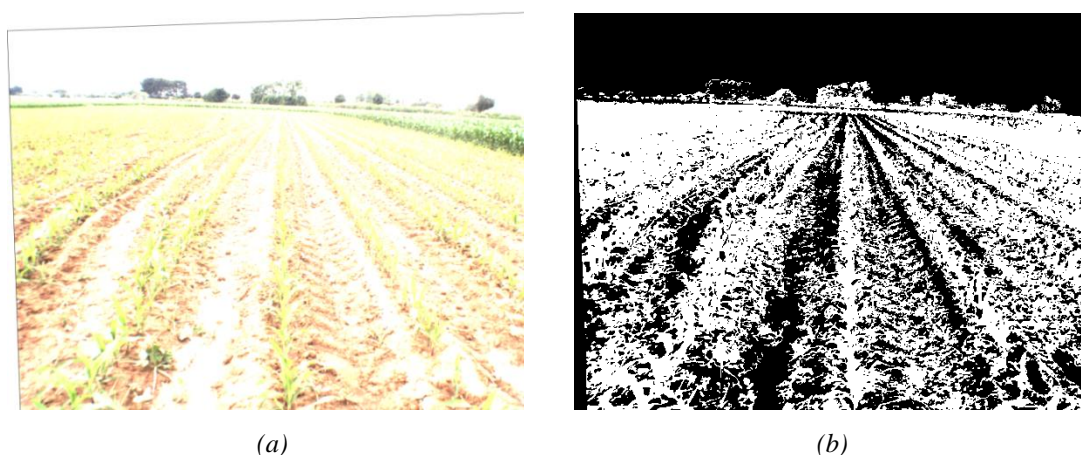
#### **4.7.2. Análisis de los factores de iluminación que afectan a la calidad de la imagen**

El segundo grupo de factores que afectan a la precisión de la imagen, en términos de calidad de la misma, basándonos en las componentes espectrales  $R$ ,  $G$  y  $B$  son los derivados de las condiciones de iluminación. Para constatar la relevancia de estos factores podemos aplicar el método de segmentación descrito en la *sección 2.3.2 "Extracción de píxeles verdes de la imagen"*, el cual obtiene un umbral de forma automática a través del método basado en la técnica de agrupamiento difuso de aprendizaje. Este umbral nos permite identificar plantas verdes del resto de la imagen proporcionando una imagen binaria. Las *figuras 4.14(a)* y *4.14(b)* muestran los resultados de aplicar el proceso a las imágenes mostradas en la *figura 4.7(a)* y *4.7(b)* respectivamente, obtenidas con dos tiempos de exposición diferentes. El tiempo de exposición de la imagen en la *figura 4.7(a)* estuvo por debajo de lo deseable, al contrario que el de la imagen obtenida en la *figura 4.7(b)*. Como puede apreciarse, el resultado en el primer caso es peor que en el segundo. La imagen binaria mostrada en la *figura 4.14(a)* contiene un exceso de segmentación en la parte de interés donde las líneas de cultivo y las malas hierbas presentan importantes vacíos en las zonas exteriores. Una posible explicación a este fenómeno es que el sensor CCD necesita tiempo suficiente para ser impactado con la reflectancia y la iluminación procedente de

los objetos de la escena. Esto es debido al hecho de que diferentes tipos de materiales producen diferentes componentes de reflectancia e iluminación que, finalmente, llegan al sensor con diferentes valores según la naturaleza del material. Cuando el tiempo de exposición no es suficiente el sensor produce este tipo de efecto. Por el contrario, si la exposición es excesiva, la incidencia de ambas componentes sobre el sensor saturan la imagen, motivo por el cual el proceso de segmentación falla. La *figura 4.15(a)* muestra una imagen saturada con su correspondiente imagen segmentada en la *figura 4.15(b)*, usando el mismo procedimiento de segmentación que en el caso anterior podemos ver cómo el resultado es no satisfactorio. Desde el punto de vista de la detección de líneas de cultivo y malas hierbas esto provoca importantes imprecisiones.



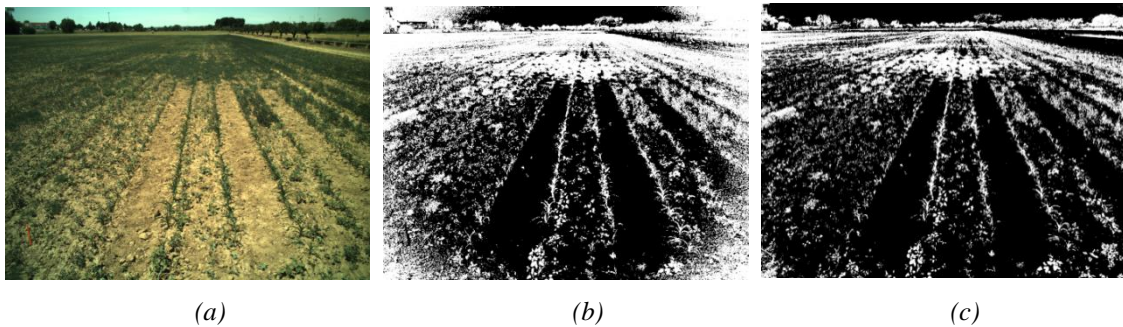
**Figura 4.14:** Imágenes segmentadas obtenidas a partir de las imágenes 4.7(a) y 4.7(b) respectivamente.



**Figura 4.15:** (a) Imagen original saturada; (b) imagen binaria después de la segmentación.

Respecto al proceso de corrección del *vignetting*, en la *figura 4.16(a)* se muestra la imagen original sin corrección, mientras que en las *figuras 4.16(b)* y *4.16(c)* se

muestran respectivamente la imagen segmentada con y sin la corrección de *vignetting*. En la *figura 4.16(b)* se puede apreciar cómo en las cuatro esquinas de la imagen aparece un exceso de píxeles representando plantas verdes, no así en la imagen de la *figura 4.16(c)* después de aplicar la corrección. Así pues, cuando no se aplica la corrección del *vignetting* aparecen imprecisiones en las imágenes, siendo tanto más acusadas cuanto más nos distanciamos del centro de la imagen, llegando a su máxima expresión en las esquinas de las imágenes. Obsérvese en la imagen *(b)* el anillo circular provocado por el efecto *vignetting*, que aparece tanto más acusado cuanto más nos alejamos del centro de la imagen, siendo de especial significado en las esquinas, donde el mismo se hace totalmente evidente.



**Figura 4.16:** (a) Imagen original; (b) Imagen binaria sin la corrección del *vignetting*, (c) Imagen binaria con la corrección del *vignetting*.



## **SISTEMA EXPERTO PARA LA IDENTIFICACIÓN DE VERDES EN IMÁGENES AGRÍCOLAS**

### **5.1 Introducción**

La arquitectura del Sistema Experto (SE) se describe en la *figura 5.1* donde aparecen identificados dos módulos principales, a saber: toma de decisiones e identificación de verdes. La imagen original, una vez capturada, se transfiere al módulo de decisión donde basándose en el análisis del histograma se toma una decisión según el contraste que ésta presente. En el caso de que la imagen contenga suficiente contraste se aplica el método combinado descrito en *Guijarro y col. 2011*, obteniendo una imagen binaria con las plantas verdes identificadas. En caso contrario, si la imagen ofrece un contraste insuficiente, se procesa primeramente aplicando un submuestreo junto con un suavizado. Después de este preprocesamiento se aplica un procedimiento automático inspirado en el agrupamiento difuso, (*Fuzzy Clustering FC*) que nos proporciona un umbral durante una fase equivalente a lo que consideramos como fase de aprendizaje en el procedimiento clásico *FC*. Una vez obtenido dicho umbral, mediante una simple regla de decisión se obtiene una imagen binaria con las plantas verdes perfectamente identificadas. A continuación realizamos una revisión sobre el estado del arte, continuando con la explicación detallada del proceso completo diseñado bajo el SE que se propone.

El presente método ha sido ideado y desarrollado dentro del proyecto RHEA (*RHEA 2014*) perteneciente al VII Programa Marco de la Unión Europea en nanociencias y nanotecnología, así como en el proyecto AGL (2011). Los resultados iniciales se publicaron en *Romeo y col. (2013b)*, *Romeo y col. (2012b)*, *Montalvo y col. (2012b)* y *Guijarro y col. (2013)*.

## 5.2. Estado del arte

Los sistemas basados en visión computacional poseen sensores que pueden utilizarse en múltiples aplicaciones. Una de ellas es en agricultura donde vehículos autónomos equipados con tales sistemas son utilizados para distinguir plantas de cultivo de malas hierbas al objeto de aplicar tratamientos específicos y selectivos *Davies y col. (1998)*. Centrándonos en los campos de maíz y cereal, uno de los tratamientos más importantes llevados a cabo es la eliminación de malas hierbas, que requieren la discriminación entre lo que son plantas de cultivo y malas hierbas. Existen diferentes métodos y estrategias en la literatura con tal finalidad, en una primera aproximación destacan entre otros los siguientes: *Onyango y Marchant (2003)*, *Tellaeché y col. (2008a-b, 2011)*, *Burgos-Artizzu y col. (2009)*, *Guijarro y col. (2011)*, *Guerrero y col., (2012a, 2013)*, *Montalvo y col. (2012a, 2013a)*. Concretamente, *López-Granados (2011)*, realiza una revisión exhaustiva de diferentes métodos donde la identificación de plantas mediante sistemas de visión artificial constituye un paso clave en el proceso. Seguidamente se presenta una revisión de métodos existentes actualmente para la identificación de verdes agrupados por categorías en función del método o proceso que aplican o realizan. Todos ellos se basan en la utilización de las componentes espectrales en el visible, por su similitud con la estrategia propuesta en este trabajo; en este sentido no se consideran otras técnicas que utilizan componentes espectrales fuera del visible, como puede ser el caso de las que se basan en el infrarrojo:

- a) *Métodos basados en el cálculo de un índice a partir de las componentes espectrales en el rango visible.* Índice de exceso de verde ExG (*Ribeiro y col. 2005*, *Woebbecke y col. 1995*), índice de exceso de rojo ExR (*Meyer y col. 1998*), índice de extracción de color de plantas CIVE (*Kataoka y col. 2003*), índice de exceso de verde menos exceso de rojo ExGR, (*Neto 2004*) e índice vegetativo VEG (*Hague y col. 2006*), este último diseñado para compensar la variación de la luz exterior. Los índices ExG, ExGR, CIVE y VEG se aplican de forma combinada en *Guijarro y col (2011)*, obteniendo un mejor resultado final que cuando se utiliza cada uno por separado. Todos estos métodos son utilizados para fijar un umbral de corte en la segmentación, es decir, para discriminar entre plantas y otros elementos existentes en el campo de cultivo, como son partes del suelo sin cubrir por plantas, ramas secas, piedras, cielo, etc.).

b) *Métodos específicos basados en el umbral, incluyendo umbrales dinámicos.*

Generalmente estos métodos asumen la existencia de dos únicos elementos a discriminar: plantas y tierra. *Reid y Searcy (1987)* establecen una función de estimación bajo la hipótesis de que las clases siguen una distribución Gaussiana. El método de *Otsu (1979)* también se aplica bajo la consideración de un problema de naturaleza bi-clase (*Ling y Ruzhitsky, 1996; Shrestha y col. 2004*). Estos algoritmos se aplican a imágenes en escala de grises, obtenidas generalmente a partir de las componentes espectrales RGB en el visible; un trabajo significativo en esta línea fue el desarrollado por *Gebhardt y col. (2006)*, quienes aplican también umbralización para la segmentación. Este método fue más tarde mejorado usando homogeneidad local y operaciones morfológicas en *Gebhardt y Kaihbauch (2007)*. *Kirk y col. (2009)*, aplicaron una combinación de verde e intensidad derivadas ambas de las bandas espectrales roja y verde y calcularon un umbral automático para un problema de dos clases asumiendo dos funciones de densidad de probabilidades Gaussianas asociadas a la tierra y a la vegetación respectivamente. Este procedimiento requiere una estimación previa de un ángulo para rotar el hipotético eje de verdes en el espacio de color RGB. *Meyer y Camargo-Neto (2008)*, aplican el umbral automático de Otsu para binarizar la imagen de grises ExG y el índice de la diferencia normalizada (NDI), estableciendo una comparación con la segmentación obtenida en ExGR llegando a la conclusión de que en este último caso, un valor de cero es suficiente para el umbral; por lo tanto, el método de Otsu no es necesario. *Guijarro y col. (2011)* y *Burgos-Artizzu y col. (2011)* aplican la media estadística de la imagen transformada obtenida con los índices de vegetación en lugar de umbrales automáticos como los de Otsu. Justifican su elección porque el método de Otsu proporciona un valor de umbral mayor que la media y por ello produce lo que técnicamente se conoce como infra-segmentación, es decir, algunas plantas no son lo suficientemente identificadas debido a que no se extrae correctamente el verde que las caracteriza. Los métodos basados en umbral no son exclusivos en aplicaciones agrícolas involucrando verdes, sino que se pueden utilizar en otras tareas agrícolas como el arado de los campos como se describe en *Moreno-Matías y Gómez-Gil (2007)*.

c) *Métodos de aprendizaje*: Meyer y col. (2004) aplican métodos de aprendizaje no supervisados, incluyendo agrupaciones difusas para segmentar regiones de interés a partir de las imágenes de grises ExR y ExG. Tian y Slaughter (1998), proponen un algoritmo de segmentación adaptativo medioambientalmente (Environmentally Adaptive Segmentation Algorithm, EASA) para detectar plantas mediante un método de aprendizaje supervisado. Ruiz-Ruiz y col. (2009) aplican el método EASA en el espacio de color HSI (Hue-Saturation-Intensity) para hacer frente a la variabilidad de la iluminación. Zheng y col. (2009) y Zheng y col. (2010), utilizan un algoritmo supervisado de desplazamiento de la media, bajo la hipótesis de que la segmentación de la vegetación del resto puede tratarse como un problema de segmentación en dos clases; la separabilidad de las clases se valida mediante una red neuronal y el discriminante lineal de Fisher respectivamente. Los espacios de color utilizados son RGB, LUV y HSI. Guerrero y col. (2012) y Tellaeché y col. (2011) aplican una técnica de aprendizaje basada en las conocidas Máquinas de Vectores Soporte para la identificación de dos clases en la imagen, una de ellas la que contiene las plantas a identificar.

La mayoría de las estrategias existentes, incluyendo las descritas anteriormente, tienen el problema de asumir que las plantas presentan un alto nivel de porcentaje de verde en relación a las tres componentes espectrales de una imagen de color en el modelo RGB. Sin embargo, las imágenes provienen del exterior donde las condiciones atmosféricas pueden provocar una gran variedad de situaciones en las que las plantas no presentan un porcentaje de la componente espectral verde tan acusado como suponen la mayoría de los métodos planteados. Esto puede deberse a una excesiva luminosidad en un día soleado o por el contrario a una homogeneidad de tonos por la escasez de sol en días nublados. Igualmente, las sombras sobre las plantas pueden hacer que éstas aparezcan negras en la imagen. La estrategia que proponemos se basa en un SE, robusto y automático específicamente diseñado para la identificación de verdes incluso en condiciones meteorológicas adversas. En la *figura 5.1* se muestra un diagrama de bloques con los distintos módulos de que consta el SE, tal y como se ha mencionado previamente, consta de los dos módulos principales, *toma de decisiones* e *identificación de verdes*, que se describen a continuación.

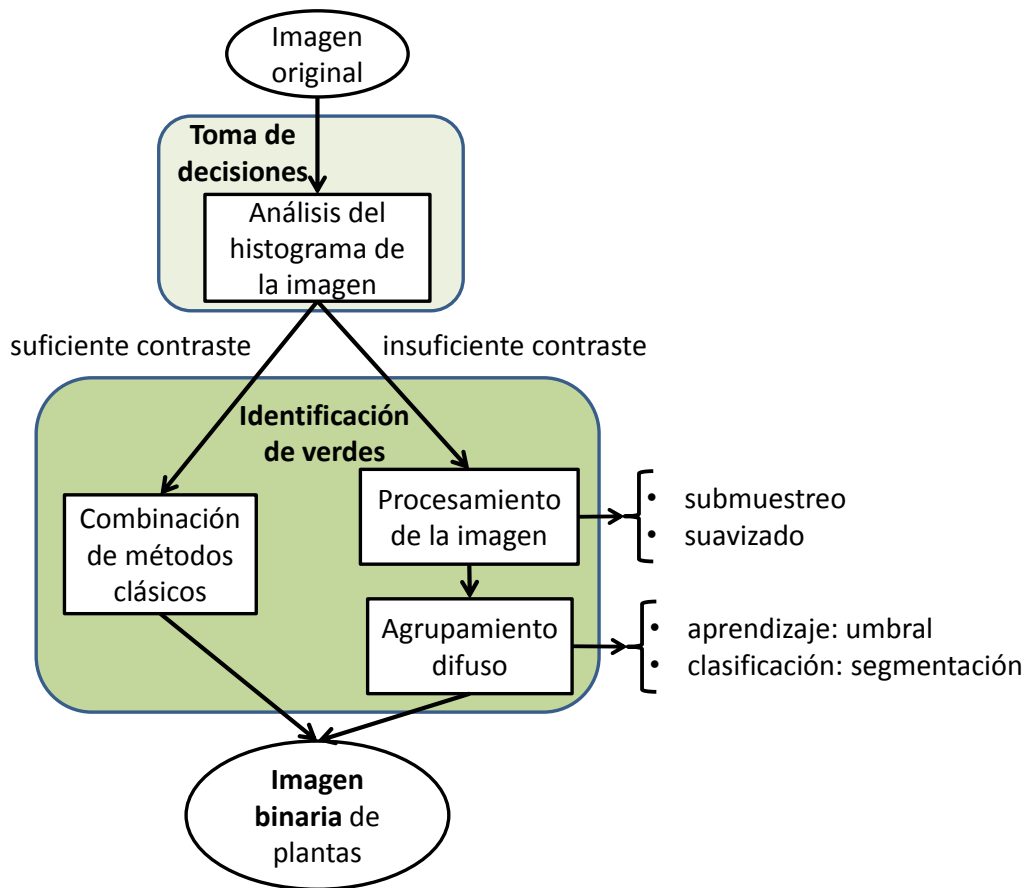


Figura 5.1: Arquitectura del sistema experto.

### 5.3. Toma de decisiones

El módulo identificado como *toma de decisiones* está diseñado con el objetivo de analizar la calidad de la imagen recibida basándose en su histograma. Después de dicho análisis se toma una decisión, de ahí su nombre, y la imagen se procesa de acuerdo con uno de los dos procesos de *identificación de verdes* presentes en el módulo. En realidad se analizan los histogramas de los tres canales espectrales RGB de la imagen. Cada histograma está definido por su correspondiente función discreta  $h(a) = n_a$ , donde  $a$  representa el nivel de gris y  $n_a$  es el número de píxeles con el nivel de gris  $a$ . En nuestra representación de los canales RGB de 8-bits por canal,  $a$  varía dentro del rango  $[0, 255]$ . Dividiendo  $h(a)$  por el número total de píxeles en la imagen obtenemos la estimación de la probabilidad para los distintos niveles de intensidad, es decir,  $p(a) = h(a)/(M \times N)$  donde  $M$  y  $N$  son las dimensiones de la imagen. El histograma proporciona estadísticas útiles que son utilizadas para determinar la calidad de la imagen. Sea  $a$  una variable aleatoria que indique niveles de gris, el momento  $n$ -ésimo de  $a$  sobre la media se define como (González y Woods 2008) se expresa en (5.1).

$$\mu_n(a) = \sum_a (a-m)^n p(a) \quad (5.1)$$

donde  $m$  es la media de  $a$ , es decir, el nivel medio de gris:

$$m = \sum_a ap(a) \quad (5.2)$$

Los parámetros estadísticos más importantes, desde el punto de vista de la calidad de la imagen, son (Holub y Ferreira 2006): *media*,  $m$ , *varianza*  $v=\mu_2(a)$ , *asimetría*  $\gamma=\mu_3(a)/\mu_2^{2/3}(a)$  y *curtosis*  $\kappa=\mu_4(a)/\mu_2^2(a)$ . La media determina el nivel promedio de intensidad. Así pues, una media baja indica que un nivel de iluminación bajo ha impactado el sensor y viceversa. Los momentos proporcionan información acerca de la distribución de valores alrededor de la media. La varianza es una medida del contraste en los niveles de gris, donde una varianza alta indica una dispersión de valores alrededor de la media y una varianza baja indica una concentración de valores alrededor de la media. La asimetría mide la falta de simetría en la distribución. Una asimetría hacia la derecha indica que el histograma muestra una cola larga hacia valores muy brillantes a la vez que una alta concentración en la parte de valores con poca intensidad (asimetría positiva). En caso contrario, cuando la cola se orienta hacia valores poco brillantes con alta concentración en la parte de valores con mucha intensidad, la asimetría es negativa. La curtosis proporciona información acerca de la zona donde aparece un pico en la distribución de intensidad, esto es en el histograma. Un valor bajo de curtosis indica zonas altas lisas alrededor de la media; sin embargo, valores altos indican picos alrededor de la media con pendientes altas. Un ejemplo ilustrativo de estos valores es el hecho de que la asimetría y curtosis son cero para distribuciones de probabilidad Gaussianas.

Una vez definidos los parámetros estadísticos indicados, la cuestión es cómo determinar la calidad de una imagen basándonos en esos valores. La respuesta viene de la experiencia en este campo. Las imágenes con suficiente contraste se consideran imágenes con calidad suficiente y viceversa. Una imagen con suficiente contraste debería presentar un histograma con valores medios en la parte central. Varianza alta, asimetría baja (tanto positiva como negativa) y un alto valor de curtosis. Por el contrario, una imagen con un contraste insuficiente presenta un histograma con valores

medios bien altos o bajos, asimetría alta (positiva o negativa) y un valor bajo de curtosis.

El siguiente paso consiste en determinar los rangos de valores posibles para los parámetros descritos anteriormente. Para ello, hemos analizado el conjunto de imágenes que se presenta y describe en la *sección 5.5 “Resultados”*, dichas imágenes fueron analizadas como imágenes con suficiente contraste o con contraste insuficiente y se calcularon sus parámetros estadísticos. Como resultado se derivaron tres reglas con cuatro premisas que enunciamos a continuación:

$$\text{Si } m_{ai} < m_i < m_{bi} \text{ y } v_i > v_{ai} \text{ y } |\gamma_i| < |\gamma_{ai}| \text{ y } \kappa_i > \kappa_{ai} \rightarrow \text{contraste suficiente} \quad (5.3)$$

en caso contrario → contraste insuficiente

donde  $i=R,G,B$ , es decir, para cada uno de los tres canales espectrales se obtienen los valores estadísticos. Los parámetros con subíndices  $a$  y  $b$  son límites superiores e inferiores que se fijarán en la *sección 5.5.1 “Módulo de toma de decisiones: identificación de parámetros”*.

## 5.4. Identificación de verdes

De acuerdo con el esquema mostrado en la *figura 5.1.*, si la imagen contiene suficiente contraste se aplican una serie de métodos ampliamente probados basados en los experimentos llevados a cabo en *Guijarro y col. (2001)*. En caso contrario se lleva a cabo una estrategia basada en *FC* para el cálculo de un umbral.

### 5.4.1. Combinación de métodos clásicos

Dada una imagen original en el espacio de color RGB, el primer paso consiste en aplicar la siguiente normalización, que habitualmente se aplica en segmentación agrícola (*Gee y col. 2008*):

$$r = \frac{R_n}{R_n + G_n + B_n}; \quad g = \frac{G_n}{R_n + G_n + B_n}; \quad b = \frac{B_n}{R_n + G_n + B_n}; \quad (5.4)$$

donde  $R$ ,  $G$  y  $B$  son las coordenadas normalizadas cuyos rangos de valores van de 0 a 1 y se calculan según la ecuación (5.5).

$$x = \frac{x}{x_{max}} \quad (5.5)$$

$x = R, G, B$  y  $x_{max} = 255$  para imágenes de color de 24-bit.

Los índices vegetativos se calculan de la siguiente manera:

$$\text{Exceso de verde, } ExG = 2g - r - b \quad (5.6)$$

$$\text{Exceso de verde menos exceso de rojo, } ExGR = ExG - 1.4r - g \quad (5.7)$$

$$\text{Índice de extrac. de veg., } CIVE = 0.441r - 0.811g + 0.385b + 18.78745 \quad (5.8)$$

$$\text{Vegetativo, } VEG = \frac{g}{r^a b^{1-a}} \quad (\text{con } a = 0,667 \text{ según Hague T. y col. 2006}) \quad (5.9)$$

Basándonos en *Guijarro y col. (2011)*, estos índices se combinan para obtener *COM* de acuerdo con la siguiente expresión:

$$COM = w_{ExG} ExG + w_{ExGR} ExGR + w_{CIVE} CIVE + w_{VEG} VEG \quad (5.10)$$

donde  $w_{ExG}$ ,  $w_{ExGR}$ ,  $w_{CIVE}$  y  $w_{VEG}$  son los pesos específicos para cada índice y representan la importancia relativa de cada uno de ellos. *Guijarro y col. (2011)* proporcionan valores para cada uno de los pesos que son los siguientes:  $w_{ExG} = 0,25$ ,  $w_{ExGR} = 0,30$ ,  $w_{CIVE} = 0,33$  y  $w_{VEG} = 0,12$ .

Los valores resultantes de la imagen combinada *COM* se restringen linealmente al rango  $[0, 1]$  y posteriormente se obtiene un umbral aplicando el método de Otsu dando como resultado una imagen binaria donde los píxeles blancos representan plantas de la imagen original con claras componentes espectrales asociadas a las plantas verdes que no presentan ninguna dificultad en su identificación. Por el contrario, los píxeles negros representan los píxeles de la imagen original que pertenecen a las plantas con cierta dificultad en su identificación, bien por presentar un color verde degradado o por estar cubiertas de lodo como consecuencia de fuertes lluvias o del propio riego del cultivo; los píxeles negros también representan tierra y otros materiales presentes en el campo. En el presente trabajo hemos optado por elegir el umbral obtenido mediante el proceso descrito en el capítulo 2 sección 2.3.2 “Extracción de píxeles verdes de la imagen”, en lugar del obtenido mediante Otsu.

### 5.4.2. Preprocesado de imagen

Cuando las imágenes no tienen suficiente contraste proponemos dos estrategias consecutivas, es decir, preprocesamiento de imágenes y *FC*. La primera de ellas se propone para reducir la resolución de la imagen, aunque podría evitarse caso de existir problemas de tiempo real. De hecho, debido a que el proceso *FC* conlleva una fase de aprendizaje asociada mediante un proceso iterativo, el número de muestras (píxeles tomados de la imagen) son determinantes en lo que se refiere al tiempo computacional. Así pues, cuanto menor es el número de píxeles a procesar menor es el tiempo de cómputo. El segundo (*FC*) es el núcleo de este proceso donde una vez realizada la fase de aprendizaje se obtiene un umbral que nos permitirá binarizar la imagen.

La imagen original es submuestreada aplicando sucesivas transformadas discretas de wavelets (*DWT*) (*Pajares y Cruz 2004*). La *DWT* se aplica a los coeficientes de aproximación en cada nivel de resolución y para cada canal espectral R, G y B en el modelo de color RGB. Al mismo tiempo que la imagen se descompone mediante la *DWT*, se aplica un proceso de suavizado para minimizar posibles ruidos o píxeles espurios debido a las condiciones medioambientales exteriores.

### 5.4.3. Agrupamiento difuso *FC*

La umbralización de la imagen mediante agrupamiento difuso consta de dos fases que describimos a continuación:

a) *Fase de aprendizaje y umbralización*

Dada la imagen submuestreada RGB, ésta contiene  $n$  píxeles almacenados en  $X$ , es decir,  $X = \{x_1, x_2, \dots, x_n\} \in R^d$ , donde  $d$  es la dimensión de los datos, donde cada dato se refiere a las tres componentes espectrales de los píxeles.

A continuación se procede de idéntica forma a la detallada en el *capítulo 2 sección 2.3.2 “Extracción de píxeles verdes de la imagen”*, para la obtención del umbral y posterior binarización de la imagen.

b) *Segmentación y clasificación*

Dada la imagen original RGB, obtenemos tres valores normalizados  $r$ ,  $g$  y  $b$  mediante la ecuación (5.5). Considerando el umbral  $T_G$  calculado en el *capítulo 2*

sección 2.3.2 “Extracción de píxeles verdes de la imagen”, la siguiente regla nos permite obtener la imagen  $W$  donde los píxeles son identificados como píxeles verdes en la imagen original:

$$W = \begin{cases} 1 & \text{si } \frac{g}{r+g+b} > T_G \\ 0 & \text{en caso contrario} \end{cases} \quad (5.11)$$

## 5.5. Resultados

Las imágenes utilizadas para el presente estudio fueron adquiridas con las siguientes cámaras digitales y en las fechas que se indican a continuación. Se tomaron tanto en los campos experimentales de la Poveda en Arganda del Rey (Madrid), como en una zona de cultivo de cereal, adaptada para simulación de maíz en la Alcarria de Trijueque (Guadalajara):

- a) HPR817 con resolución 600 x 800 en marzo/abril/mayo 2007 (330 imágenes) en campos de maíz y cereal.
- b) Basler 17FC 1400 conectada a un portátil mediante FireWire IEEE 1394 con resolución 1280x960 en marzo 2011 (320 imágenes) en campos de maíz.
- c) Kodak EasyShare M380 con resolución 1280 x 960 en marzo 2011 en campos de cereal y en mayo 2012 en campos de maíz (210 imágenes).
- d) Canon EOS 400D con resolución de 1944 x 2592 en abril/mayo 2011 (350 imágenes) en campos de maíz y cereal.
- e) SVS-VISTEK con conexión GigaEthernet y resolución 2336 x 1752 equipada con un filtro de corte de radiación Schneider (*Schneider 2014b*) UV-IR 486 en marzo/abril/mayo de 2012 (340 imágenes) en campos de cereal y maíz.

En total 1550 imágenes fueron analizadas. Las imágenes fueron tomadas en diferentes días bajo condiciones medioambientales variadas, a distintas horas del día y con

distintas orientaciones de la cámara con respecto a la principal fuente de iluminación que es el sol.

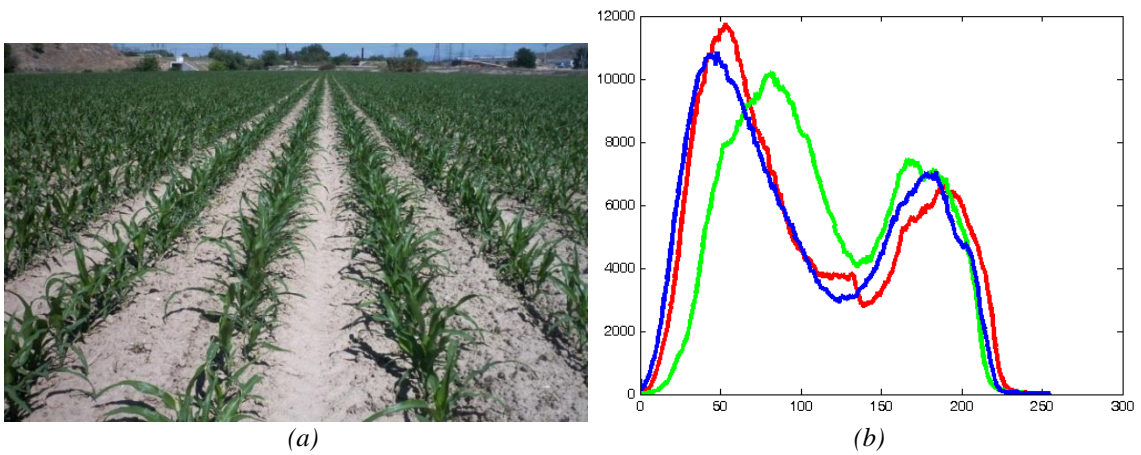
Las imágenes fueron tomadas con proyección perspectiva en campos de cereal y maíz reales en diferentes áreas geográficas, tal como se ha indicado previamente, dichos campos contenían los elementos naturales propios de los mismos, tales como piedras, ramas secas y demás residuos naturales aparte de los cultivos reseñados y las malas hierbas características de cada lugar. Estas imágenes se almacenaron en formato BMP y JPEG, como bien es sabido el primero conserva las propiedades de las imágenes ya que carece de compresión, a diferencia del JPEG donde debido a la compresión a veces aparecen lo que técnicamente se conoce como “artefacto”, que afecta a la calidad de la imagen.

El sistema experto con el que se han procesado las imágenes anteriores fue implementado en Matlab R2009b (*The Mathworks 2014*) y las imágenes fueron procesadas utilizando la Toolbox Image Processing, específica para el tratamiento y procesamiento de imágenes.

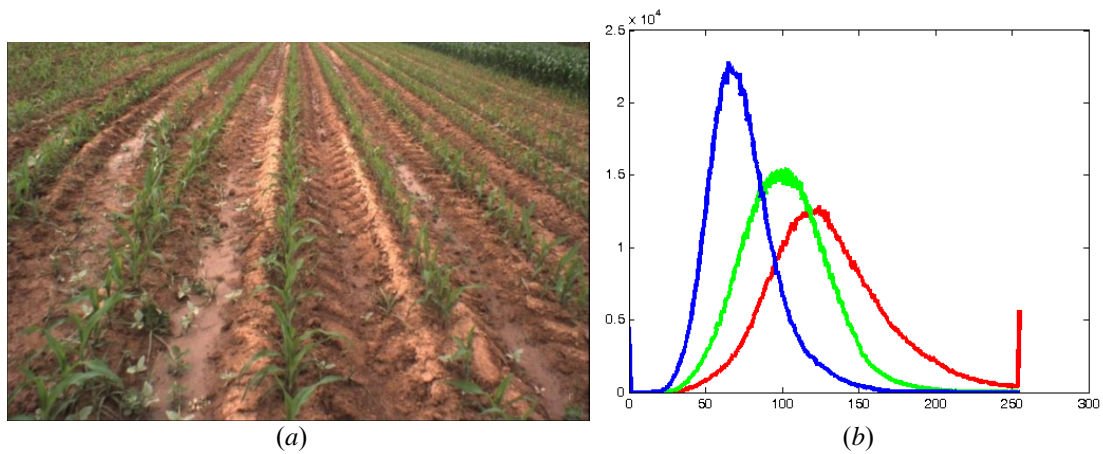
### **5.5.1. Módulo de toma de decisiones: identificación de parámetros**

Del conjunto de imágenes disponibles, seleccionamos de forma aleatoria un 5% de las imágenes tomadas con cada cámara. De acuerdo con el módulo de *identificación de verdes*, para cada imagen obtuvimos dos imágenes binarias, una aplicando la combinación de métodos clásicos según la ecuación (5.10) y un segundo aplicando el preprocesamiento y el agrupamiento difuso, descrito en las *secciones 5.4.2 y 5.4.3* respectivamente. Las imágenes fueron visualmente analizadas por un experto para determinar los mejores resultados en cuanto a partes verdes obtenidas se refiere comparándolas con las imágenes originales.

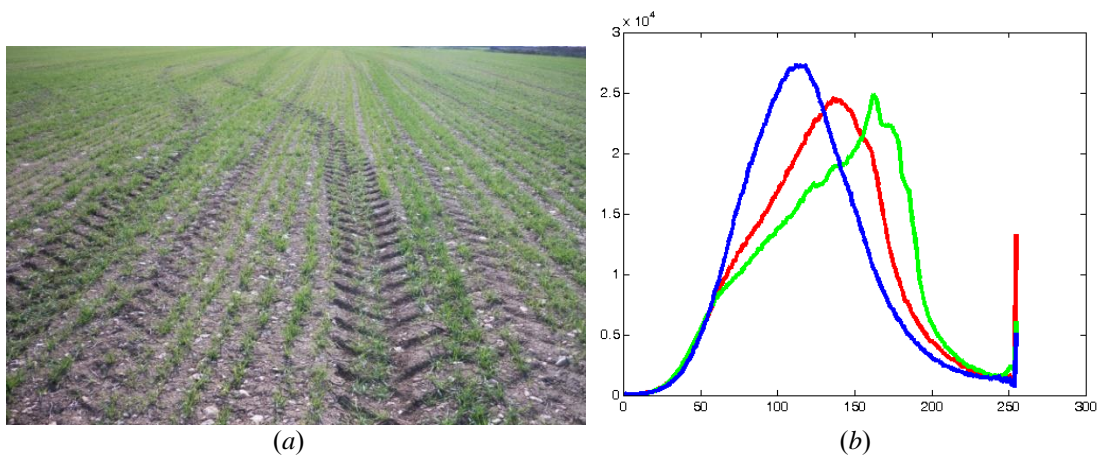
Las *figuras 5.2 a 5.4(a)* muestran tres imágenes representativas del conjunto de imágenes disponibles en las que la mejor binarización se ha obtenido mediante los métodos clásicos. Por el contrario las imágenes mostradas en las *figuras 5.5a 5.7(a)* muestran tres imágenes en las que la mejor binarización se ha obtenido mediante el agrupamiento difuso, es decir, son imágenes con insuficiente contraste. La serie (*b*) de las *figuras 5.2 a 5.7* muestran los correspondientes histogramas para los tres canales espectrales R, G y B de cada una tales imágenes.



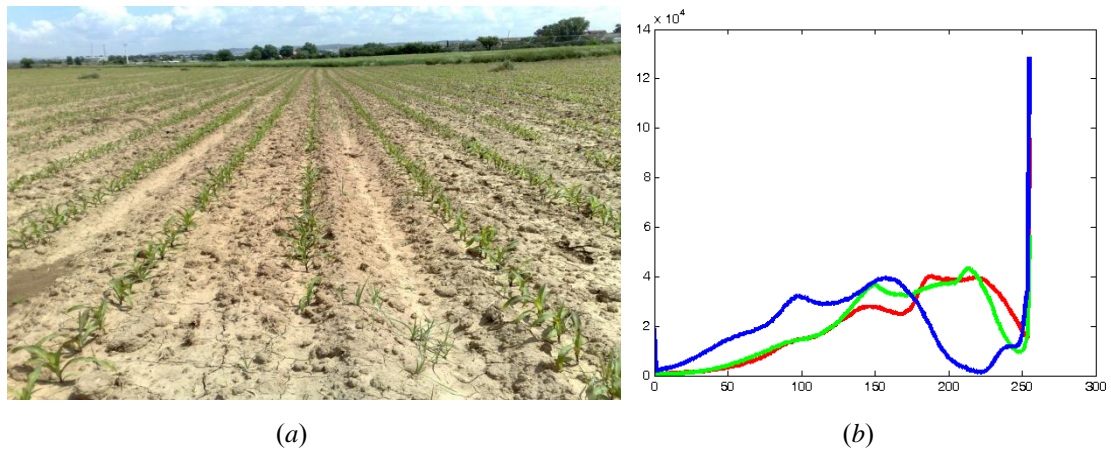
**Figura 5.2:** (a) Imagen original (maíz) obtenida con una cámara Kodak EasyShare M380; (b) histograma de los tres canales espectrales.



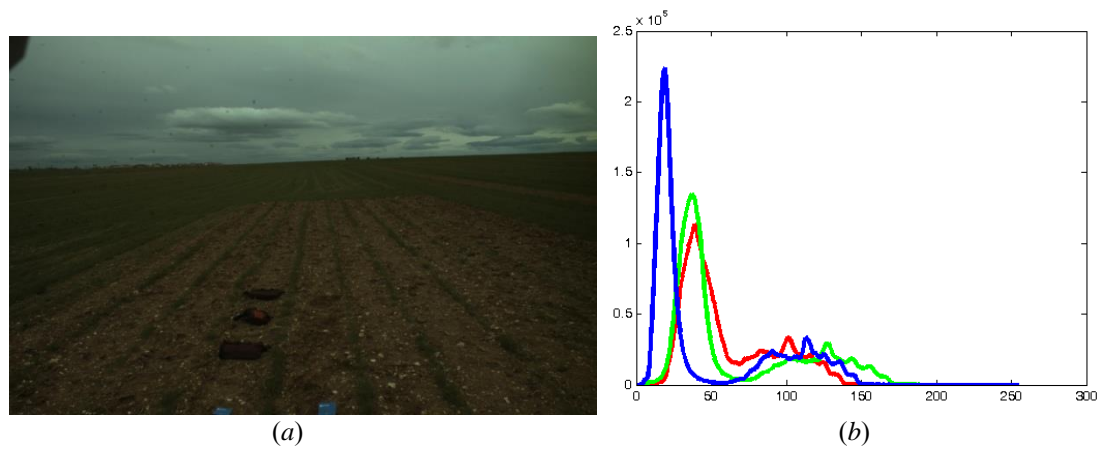
**Figura 5.3:** (a) Imagen original (maíz) obtenida con una cámara Basler 17FC 1400; (b) Histograma de los tres canales espectrales.



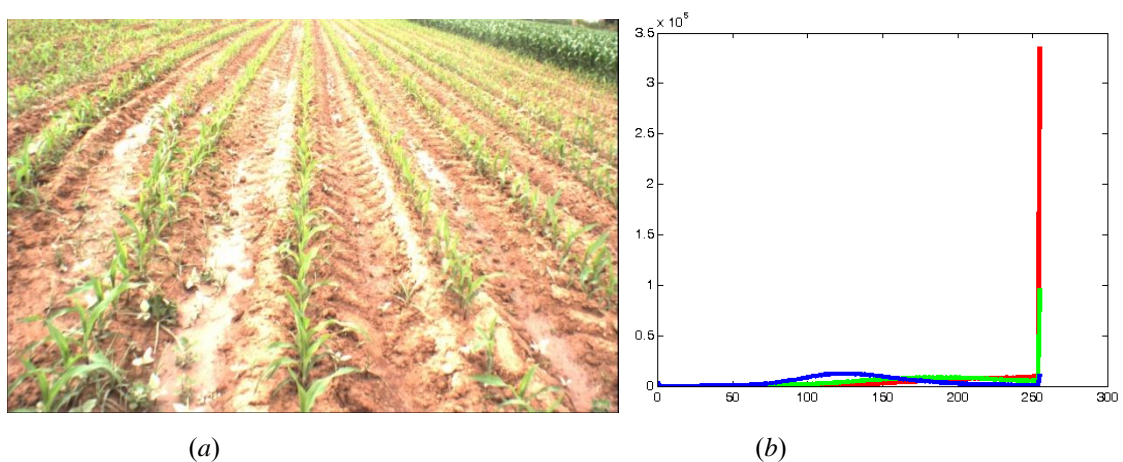
**Figura 5.4:** (a) Imagen original (cereal) obtenida con una cámara HPR817; (b) histograma de los tres canales espectrales.



**Figura 5.5:** (a) Imagen original (maíz) obtenida con una cámara Canon EOS 400D; (b) Histograma de los tres canales espectrales.



**Figura 5.6:** (a) Imagen original (trigo en líneas de cultivo anchas simulando maíz) obtenida con una cámara VISTEK SVS4; (b) Histograma de los tres canales espectrales.



**Figura 5.7:** (a) Imagen original (maíz) obtenida con una cámara Basler 17FC 1400 SVS4; (b) Histograma de los tres canales espectrales.

Las *figuras 5.2, 5.3 y 5.4* muestran las imágenes con sus correspondientes histogramas que cubren la totalidad del rango de valores con los máximos localizados en la zona central. Por el contrario, los histogramas de las *figuras 5.5, 5.6 y 5.7* están inclinados hacia la derecha o izquierda. Esto representa un hallazgo importante para la toma de decisiones. La única cuestión es cuantificar esta observación de alguna manera.

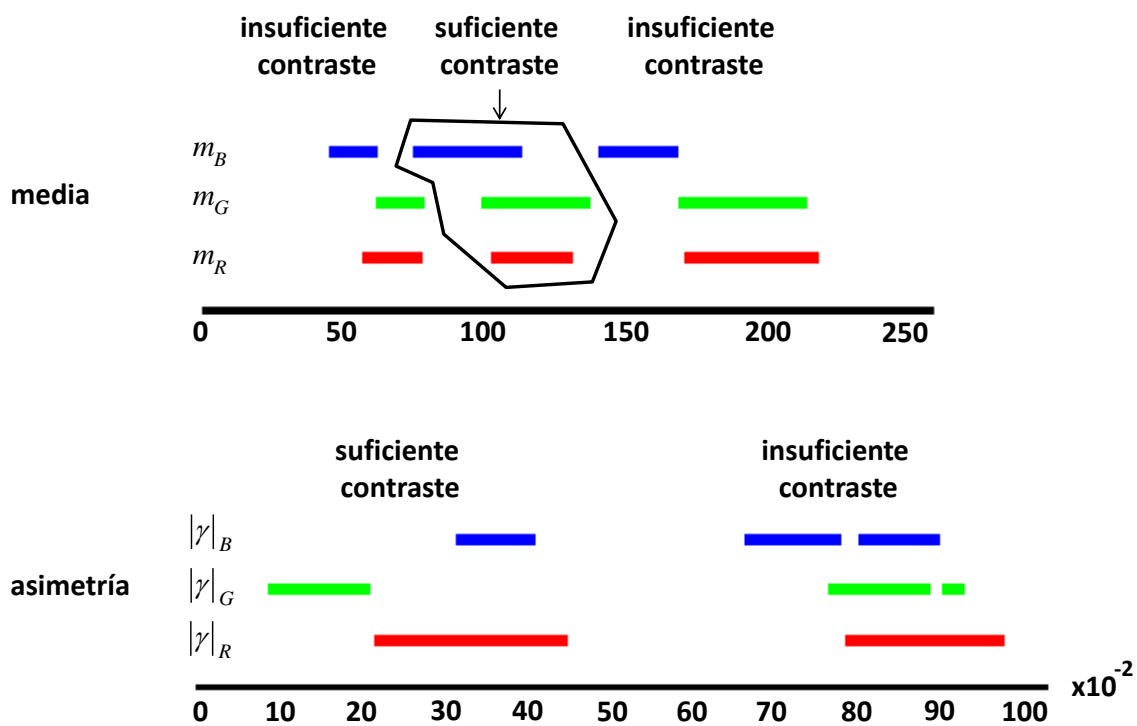
La *tabla 5.1* muestra los valores promedio de los parámetros obtenidos a partir de los correspondientes histogramas de las imágenes analizadas de acuerdo con la mejor imagen binaria obtenida. Los valores máximos y mínimos de cada canal espectral R, G y B están identificados por la media  $m$ , la varianza  $\nu$ , la asimetría  $\gamma$  y la curtosis  $\kappa$ . Gracias a estos valores podemos distinguir entre imágenes con suficiente contraste o con insuficiente contraste. Además, de estos últimos podemos distinguir entre imágenes con el histograma inclinado a la derecha o a la izquierda. Esto es debido a que la media varía de forma distinta, lo cual queda también indicado por los valores de la asimetría.

	Contraste suficiente		Contraste insuficiente			
			Inclinado a la izquierda		Inclinado a la derecha	
	Mínimo (R,G,B)	Máximo (R,G,B)	Mínimo (R,G,B)	Máximo (R,G,B)	Mínimo (R,G,B)	Máximo (R,G,B)
$m$	(106, 102, 78)	(130, 136, 112)	(60, 65, 49)	(75, 81, 60)	(174, 168, 144)	(217, 213, 167)
$\nu \times 10^3$	(1,51 0,83 0,52)	(3,73 2,64 3,65)	(0,91 1,83 1,85)	(2,91 3,55 3,89)	(1,83 2,20 1,72)	(2,72 2,53 3,33)
$ \gamma $	(0,22 0,10 0,33)	(0,45 0,21 0,41)	(0,85 0,89 0,82)	(0,89 0,93 0,90)	(0,79 0,78 0,68)	(0,98 0,87 0,78)
$\kappa$	(1,75 1,84 1,72)	(3,69 3,73 4,50)	(2,46 2,26 2,10)	(4,45 3,98 4,60)	(2,77 2,74 2,69)	(5,42 3,08 3,66)

**Tabla 5.1:** Parámetros del histograma (media, varianza, asimetría y curtosis) identificando los valores máximo y mínimo para imágenes con suficiente contraste en los tres canales R, G y B.

A partir de los resultados de la *tabla 5.1* se pueden extraer dos conclusiones importantes. Existen dos parámetros que son realmente discriminantes ( $m$  y  $\gamma$ ) y dos que no lo son ( $\nu$  y  $\kappa$ ). Si observamos el canal rojo podemos ver que la media para imágenes con contraste suficiente varía en el rango [106, 130] y para imágenes con insuficiente contraste aparecen dos intervalos: uno para los inclinados a la izquierda [60, 75] y otro para los inclinados a la derecha [174, 217]. Estos intervalos no se superponen y mantienen un amplio margen entre las correspondientes medias. Este razonamiento

puede extenderse igualmente para los canales verde, G y azul, B, en los que aparece un comportamiento similar, es decir, [102, 136] y [65, 81]-[168, 213] para el verde y [78, 112] y [49, 60]-[144, 167] para el azul. Con respecto a la asimetría y el canal rojo podemos ver cómo imágenes con suficiente contraste tienen un rango de valores en el intervalo [0,22, 0,45] e imágenes con insuficiente contraste varían entre [0,85, 0,89] y [0,79, 0,98], es decir, aparece una clara superposición de valores en estos intervalos. Un comportamiento similar se produce en el canal verde con los valores en los rangos [0,10, 0,21] y [0,89, 0,93]-[0,78, 0,87] y en el canal azul con [0,33, 0,41] y [0,82, 0,90]-[0,68, 0,78]. En la *figura 5.8* se puede apreciar más claramente esta situación donde podemos observar que no existe superposición de valores en los valores promediados de la media y de la asimetría. Cada canal está identificado con su correspondiente subíndice R, G y B.



*Figura 5.8:* Representación de los valores medios y oblicuidad para los tres canales espectrales R, G y B.

El comportamiento de la media y de la asimetría no se observa con la varianza y curtosis donde se produce superposición de valores y por tanto no sirven como parámetros discriminitorios. Por tanto, fijamos los valores de la media y de la asimetría de la siguiente forma:  $m_{aR} = 106$ ;  $m_{aG} = 102$ ;  $m_{aB} = 78$ ;  $m_{bR} = 130$ ;  $m_{bG} = 136$ ;  $m_{bB} = 112$ , podemos asumir un cierto margen de tolerancia porque hay distancia suficiente

para evitar la superposición en el rango de valores. Para la asimetría los valores apropiados serían entre 0.50 y 0.60, es decir,  $\gamma_{aR} = \gamma_{aG} = \gamma_{aB} = 0,55$ . Por las razones expresadas anteriormente, tanto la varianza como la curtosis no son tenidas en cuenta como parámetros discriminantes, razón por la cual y se eliminan de la ecuación (5.3).

### 5.5.2. Módulo de identificación de verde

Una vez que los parámetros mencionados previamente han sido establecidos, seleccionamos aleatoriamente un 15% de las imágenes disponibles, es decir, 232 imágenes. La mitad de ellas con contraste suficiente y la otra mitad con contraste insuficiente. Todas ellas distintas de las utilizadas para la identificación de los citados parámetros. Los resultados de esta estrategia combinada y también de la basada en *FC* se comparan con las imágenes consideradas como realmente válidas, denominadas técnicamente como “ground-truth”, las cuales son construidas según el procedimiento descrito en *Guijarro y col. (2011)*, resumido brevemente a continuación:

1. De la imagen en el espacio de color RGB extraemos todos los píxeles cuya componente verde *G*, es mayor que la componente roja, *R* y la componente azul, *B* simultáneamente al menos en un 10%. Esto nos permite obtener una parte importante de plantas verdes que se etiquetan como blanco.
2. Igualmente de la imagen original extraemos los píxeles donde el rojo y el azul son dominantes, es decir, superan en al menos un 10% a las otras dos componentes. Esto nos permite extraer el rojo y el azul con un cierto grado de confianza, etiquetándolos como negro.
3. Los píxeles que aún no han sido etiquetados, los visualizamos y etiquetamos manualmente comparándolos con la imagen original. Sin embargo, cuando las imágenes son complejas como la de la *figura 5.6(a)* o extremadamente complejas como la de la *figura 5.10(a)* seleccionamos partes que contienen plantas verdes y otros elementos que son fácilmente distinguibles.

### 5.5.3. Combinación de métodos clásicos

Hemos aplicado directamente los parámetros propuestos en *Guijarro y col. (2011)*, obteniendo resultados similares.

#### 5.5.4. Preprocesamiento de imagen

Aplicamos sucesivas transformadas de wavelets discretas (*DWT*). Como hemos mencionado previamente, este procedimiento implica un submuestreo y un proceso de filtrado. Las imágenes con la máxima resolución son las proporcionadas por la cámara Canon EOS 400D (1944 x 2592). Comprobamos que reducciones de hasta 1/16 son suficientes para conseguir resultados coherentes en términos de efectividad, comparándolos con las correspondientes imágenes “ground-truth”, además de reducir considerablemente el tiempo de cálculo durante este proceso y consecuentemente durante el proceso de agrupamiento difuso debido a un número menor de muestras a procesar. Por tanto, una resolución de 121 x 162 supone la obtención de 19.602 muestras para nuestros experimentos. Para imágenes originales con resoluciones distintas a la mencionada, la descomposición en *DWT* debe ser tal que se alcance un número similar de muestras. Respecto a la familia de wavelets utilizada hemos podido comprobar que no es relevante y por ello hemos utilizado la que implica menor coste computacional. La transformada de Haar cumple con nuestros requisitos porque su filtro de naturaleza *paso bajo* contiene un menor número de coeficientes y por lo tanto el número de operaciones para el filtrado es menor.

#### 5.5.5. Agrupamiento difuso *FC*

Como ya hemos indicado previamente, este proceso comienza con los centros de clases ya inicializados, a continuación aplicamos sucesivas iteraciones para estimar los centros de clases definitivos, ecuaciones (2.2) y (2.3), hasta conseguir la convergencia en base al parámetro de tolerancia  $\varepsilon$  o se alcance el número máximo de iteraciones,  $t_{max}$ .

Para fijar estos parámetros hemos seleccionado 10 de las imágenes disponibles identificadas como imágenes con contraste insuficiente con su correspondiente “ground-truth”. Inicialmente fijamos  $t_{max} = 100$ , lo cual representa un número relativamente elevado. Variamos  $\varepsilon$  de  $10^{-2}$  a  $10^{-5}$  en pasos de  $5 \times 10^{-5}$ , y una vez que hemos obtenido los resultados apropiados, anotamos el número de iteraciones  $t$ , y el valor de  $\varepsilon$  para cada imagen.

La *tabla 5.2* muestra estos valores para las diez imágenes utilizadas.

Nº Imagen	1	2	3	4	5	6	7	8	9	10
Iteraciones( $t$ )	3	8	12	2	3	1	6	7	3	2
$\varepsilon \times 10^{-5}$	6	11	14	77	73	24	18	10	12	15

*Tabla 5.2: Iteraciones y valores de  $\varepsilon$  para diez imágenes analizadas.*

Con los valores de la *tabla 5.2* fijamos  $t_{max}=14$  y  $\varepsilon=10^{-5}$ , lo cual nos garantiza un cierto grado de confianza ya que son los valores más restrictivos. Estos valores son los que vamos a utilizar en posteriores análisis.

### 5.5.6. Método general: análisis cuantitativo

El sistema experto SE propuesto es ahora analizado utilizando 222 imágenes con sus correspondientes “ground-truth”. Las imágenes utilizadas para fijar los valores de  $t_{max}$  y de  $\varepsilon$  no han sido utilizadas en este análisis.

Este conjunto de imágenes han sido previamente clasificadas como imágenes con contraste suficiente (108) e imágenes con contraste insuficiente (114) por un experto. Para evaluar los resultados de la clasificación se ha utilizado uno de los índices más clásicos para tal fin el cual es el conocido como *porcentaje de clasificación correcto* (Correct Classification Percentage PCC, *Rosin 2003*). La nomenclatura utilizada para el cómputo de este índice es la siguiente:

- *SC*: Número de imágenes identificadas correctamente con contraste suficiente.
- *IC*: Número de imágenes identificadas correctamente con contraste insuficiente.
- *SCE*: Número de imágenes identificadas con contraste suficiente siendo con contraste insuficiente, es decir, identificadas erróneamente.
- *ICE*: Número de imágenes identificadas con contraste insuficiente siendo con contraste suficiente, es decir, identificadas erróneamente.

Así pues calculamos el índice *PCC* de la siguiente forma:

$$PCC = \frac{SC + IC}{SC + IC + SCE + ICE} \quad (5.12)$$

Por último, de acuerdo con la regla de decisión de la ecuación (5.3), eliminando la varianza y la curtosis por las razones explicadas previamente, obtenemos  $SC = 101$ ,  $IC = 109$ ,  $SCE = 7$  e  $ICE = 5$  lo cual es un resultado muy aceptable.

La combinación de métodos clásicos se prueba con las imágenes  $IC$  y  $SCE$  obteniendo un 91,0 % de aciertos. Este porcentaje se obtiene con las correspondientes imágenes “ground-truth” de las imágenes utilizadas para la clasificación. Los resultados obtenidos en nuestros experimentos son muy similares a los obtenidos en *Guijarro y col. (2011)*.

La validez de nuestro método  $FC$  se verifica con las imágenes  $IC$  y las  $SCE$  obteniendo un porcentaje de éxito del 89,9%.

La *tabla 5.3* muestra el promedio de tiempo de cálculo utilizado por la CPU en cada uno de los procesos de nuestro SE para el conjunto de imágenes utilizadas. El valor en  $FC$ , es un promedio de las iteraciones realizadas, es decir, una media del tiempo consumido en cada iteración.

Basándonos en los resultados de la *tabla 5.3*, cuando una imagen se identifica como imagen con contraste suficiente, el tiempo total de CPU utilizado es 0,91s; cuando la imagen es identificada como con contraste insuficiente y en el peor de los casos, es decir con  $t_{max} = 14$ , el tiempo total de CPU utilizado es 0,6s. Esto significa que el tiempo total de todo el proceso en este caso sería de 1,62s.

	<b>Análisis de histograma</b>	<b>COM</b>	<b>DWT</b>	<b>FC (por iteración)</b>
<b>Tiempo CPU (s)</b>	0,19	0,72	1,01	0,03

*Tabla 5.3: Tiempos utilizados por la CPU (en segundos) en los procesos implicados en el sistema experto.*

Teniendo en cuenta que los tractores actuales, donde se instalan los sistemas de visión mediante los que se obtienen las distintas imágenes, cuando están realizando tareas de identificación de plantas navegan a una velocidad de entre 4 y 6 km/h (1,1 – 1,7 m/s) y que los sistemas de visión actuales pueden cubrir áreas de 4 metros de longitud con resolución suficiente, disponemos de un margen de entre 3,6 y 2,35 segundos para procesar la imagen. Esto nos lleva a la conclusión que los tiempos requeridos por nuestro SE están dentro de lo razonable.

### 5.5.7. Método general: análisis cualitativo

Las *figuras 5.9(a)-5.11(a)* muestran imágenes originales representativas del conjunto de imágenes disponibles de campos de maíz. En la parte *(b)* de dichas figuras pueden verse sus correspondientes histogramas. Los resultados y las conclusiones obtenidas de este análisis son válidos para el conjunto completo de imágenes probadas. En la *figura 5.10(a)* podemos ver una zona limitada del campo con diferentes densidades de malas hierbas entre las líneas de cultivo. Alternativamente aparecen zonas con densidades altas y bajas que fueron expresamente preparadas por los agricultores con el objetivo de poner a prueba nuestro método. La *tabla 5.4* muestra los parámetros más relevantes (media  $m$ , asimetría  $\gamma$ , centro de la clase representando los verdes  $v_I$  y umbral  $T_G$ ) que utilizamos para tomar la decisión sobre qué proceso aplicar, es decir, *COM* o agrupamiento difuso *FC*. Para cada imagen anotamos el tipo de cámara utilizada.

	<b>Figura 5.9. (Kodak)</b>	<b>Figura 5.10. (SVS-VISTEK)</b>	<b>Figura 5.11. (Basler)</b>
$(m_R, m_G, m_B)$	(116, 124, 105)	(52, 53, 33)	(236, 230, 187)
$(\gamma_R, \gamma_G, \gamma_B)$	(0,32 0,17 0,38)	(0,68 0,89 1,65)	(-2,74 -2,13 -1,43)
$v_I \equiv \{v_{IR}, v_{IG}, v_{IB}\}$	(71,2 91,3 57,1)	(32,0 34,6 19,9)	(176,8 155,0 107,2)
$T_G$	0,42	0,40	0,35

**Tabla 5.4:** Valores de los parámetros del histograma (media, asimetría, centro de clúster  $v_I$  y umbral  $T$ ) para las imágenes mostradas en las figuras 9, 10 y 11.

A partir de los resultados mostrados en la *tabla 5.4* podemos ver, de acuerdo con la regla de decisión ecuación (5.11), que la imagen de la *figura 5.9(a)* está identificada como imagen con contraste suficiente. Por lo tanto debería ser procesada aplicando el procedimiento *COM* según la ecuación (5.10).

Por otro lado, las imágenes de las *figuras 5.10(a)* y *5.11(a)* han sido identificadas como imágenes con contraste insuficiente; esto significa que deberían ser procesadas con el procedimiento de agrupamiento difuso *FC*. De acuerdo con el criterio del experto, las tres decisiones tomadas para estas imágenes son correctas y los histogramas confirman este hecho.

Un aspecto importante a tener en cuenta es la variabilidad del umbral  $T_G$ , que en imágenes con contraste insuficiente varía de 0,34 a 0,52 dependiendo del tipo de imagen. Por lo tanto, establecer un valor del umbral como fijo y universal no es posible de forma general. Este hecho aporta una contribución importante de este método, es decir nuestro SE, proporciona un umbral dinámico, adaptado a la calidad de la imagen, mediante el análisis de su contraste. Además, se puede extraer información adicional del histograma. Este es el caso de la imagen mostrada en la *figura 5.11(a)* y su histograma en *(b)*, donde uno puede darse cuenta fácilmente que la imagen presenta sobresaturación.

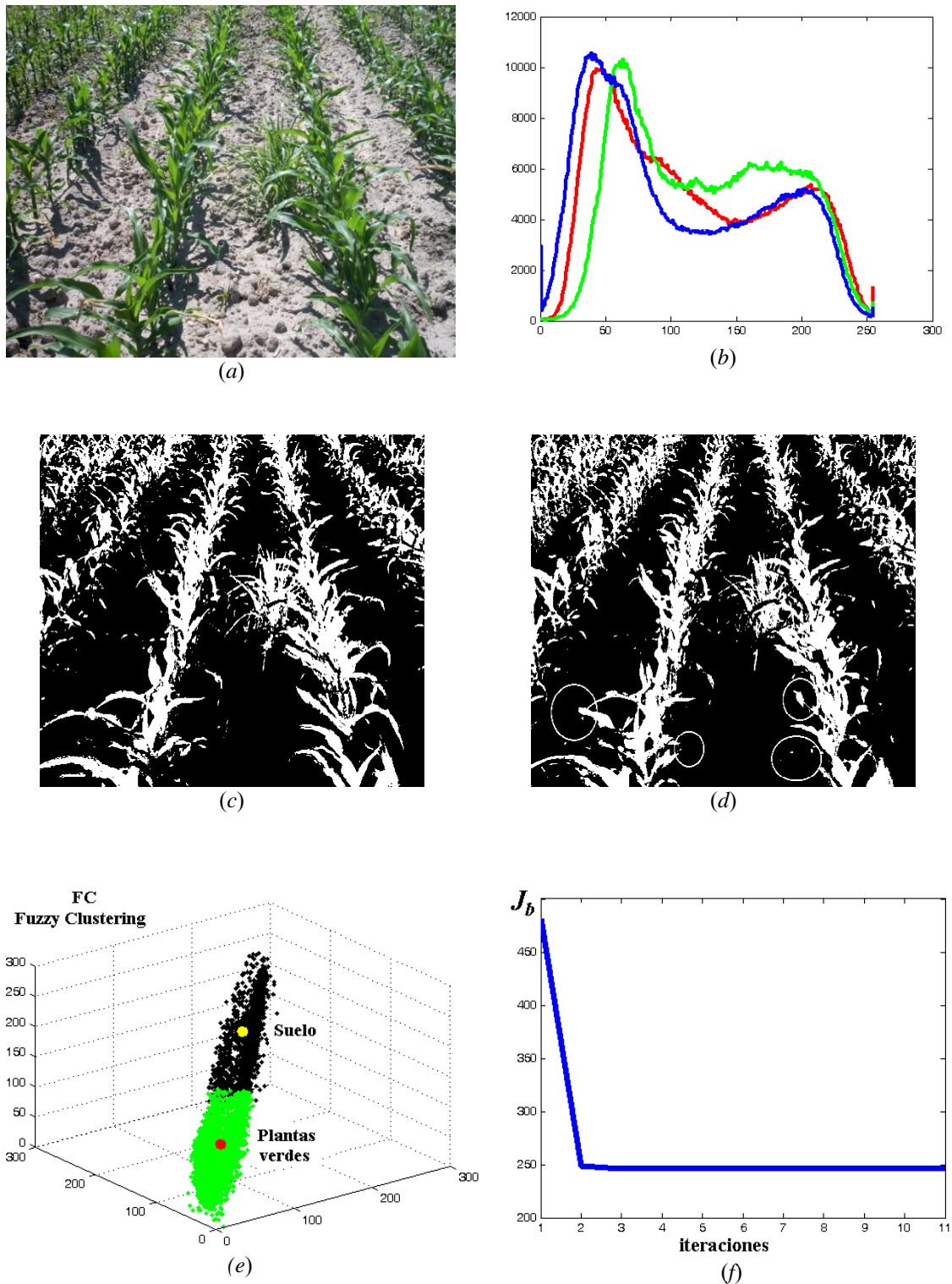
Aunque el método de agrupación difusa propuesto proporciona resultados que permiten trabajar bajo condiciones adversas críticas, la mejor opción sería controlar la entrada de iluminación al sensor CCD de la cámara, bien controlando la apertura del diafragma o bien teniendo acceso al tiempo de exposición.

A efectos exclusivamente comparativos de los resultados obtenidos, hemos procesado las tres imágenes tanto con *COM* como con el agrupamiento difuso *FC*, obteniendo respectivamente las correspondientes imágenes binarias mostradas en las diferentes *figuras 5.9(a)* y *5.11 (c)-(d)*.

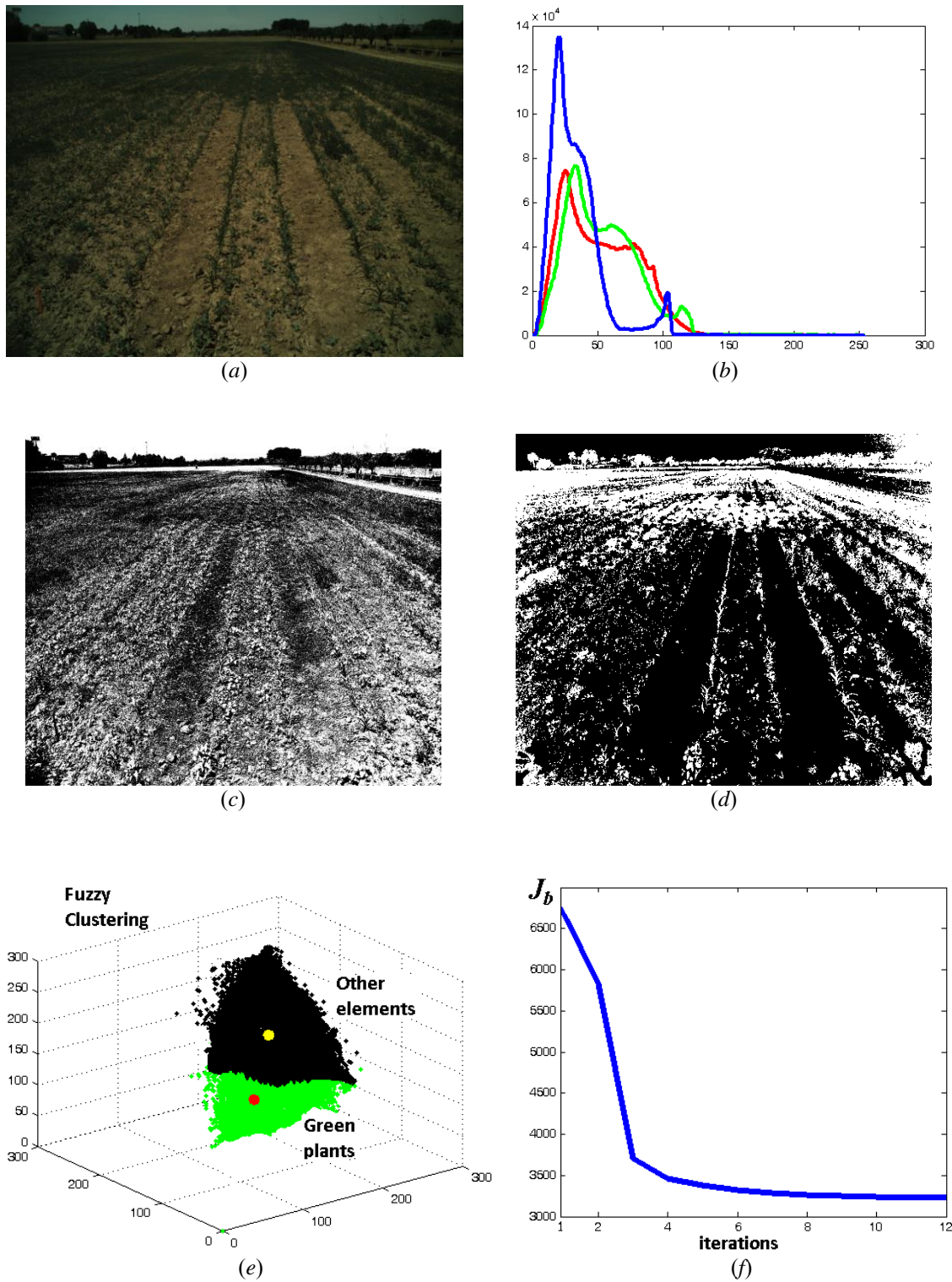
Igualmente y con el mismo objetivo en las imágenes de las figuras etiquetadas con *(e)*, se muestran los resultados obtenidos mediante la clasificación de las muestras con los centros de clases identificados como vegetación y el resto como otros elementos (tierra, cielo, etc.).

Finalmente, las figuras etiquetadas bajo la serie *(f)* muestran la variación de la función criterio definida en la ecuación (5.10) utilizada por *FC*, frente al número de iteraciones.

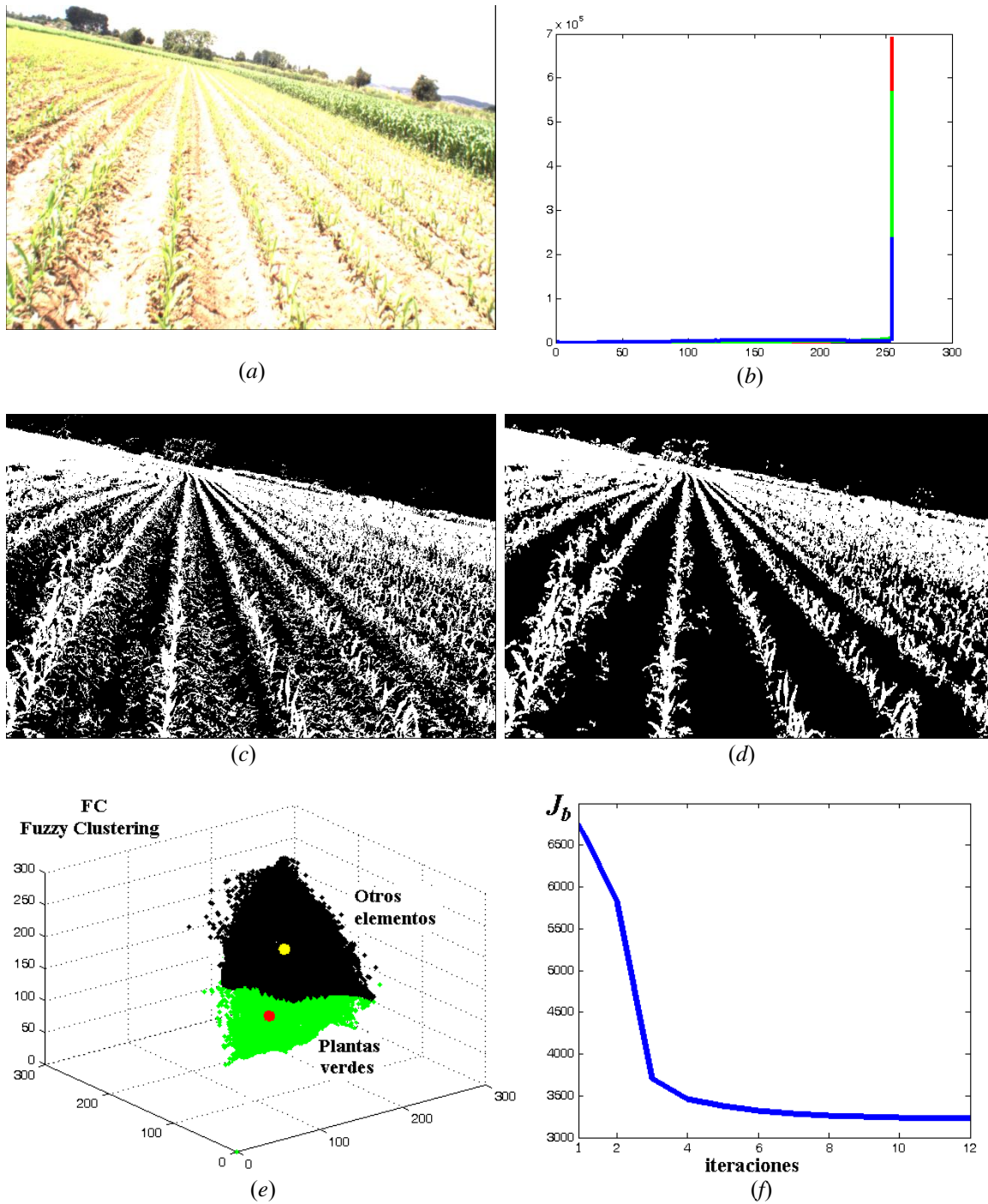
Es importante tener en cuentas una vez más, que las imágenes con contraste suficiente no son tenidas en cuenta por el método de agrupamiento difuso *FC* y por tanto no procesadas por éste, mientras que las imágenes con contraste insuficiente no son procesadas con *COM*.



**Figura 5.9:** (a) Imagen original representativa del conjunto de imágenes con suficiente contraste, (b) Histograma de la imagen para los tres canales espectrales RGB, (c) Imagen binaria obtenida con COM, (d) Imagen binaria obtenida con FC, (e) Distribución de centros y clases obtenidos con la agrupación difusa FC y (f) Variación de la función del criterio frente al número de iteraciones.



**Figura 5.10:** (a) Imagen original representativa del conjunto de imágenes con contraste insuficiente, (b) Histograma de la imagen para los tres canales espectrales RGB, (c) Imagen binaria obtenida con COM, (d) Imagen binaria obtenida con FC, (e) Distribución de centros y clases obtenidos con la agrupación difusa FC y (f) Variación de la función del criterio frente al número de iteraciones.



**Figura 5.11:** (a) Imagen original representativa del conjunto de imágenes con contraste insuficiente, (b) Histograma de la imagen para los tres canales espectrales RGB, (c) Imagen binaria obtenida con COM, (d) Imagen binaria obtenida con FC, (e) Distribución de centros y clases obtenidos con la agrupación difusa FC y (f) Variación de la función del criterio frente al número de iteraciones.

### 5.5.8. Análisis de imágenes binarias

Comparando las imágenes binarias obtenidas mediante COM y mediante agrupamiento difuso FC, podemos fácilmente verificar la validez del procedimiento de decisión utilizado. La calidad de la imagen binaria en la figura 5.9(c) es mejor que la de

la *figura 5.9(d)* donde las irregularidades más importantes están marcadas con círculos, es decir, plantas que no han sido identificadas. Esto significa que *COM* supera en resultados al agrupamiento difuso *FC* como era de esperar. Por el contrario, la calidad de las imágenes en las *figuras 5.10(d)* y *5.11(d)* es claramente superior a la obtenida en las *figuras 5.10(c)* y *5.11(c)*. De aquí se deduce que el agrupamiento difuso *FC* es el método apropiado para imágenes con contraste insuficiente en lugar de *COM*. Estos resultados confirman de una forma cualitativa que el *SE* propuesto es apropiado para identificar verdes (vegetación) en imágenes agrícolas tomadas incluso bajo condiciones medioambientales adversas.

### 5.5.9. Análisis del agrupamiento difuso *FC*

Las *figuras 5.9-5.11(e)* muestran el establecimiento de clases para las plantas verdes y el resto de elementos en la imagen, principalmente tierra, cielo, árboles y algún edificio. Como podemos ver en estas tres imágenes, las muestras pertenecientes a cada clase aparecen agrupadas. Quizá una mayor compacidad puede verse en imágenes con contraste insuficiente, *figuras 5.10(e)* y *5.11(e)*, lo cual es una consecuencia directa del histograma y de su concentración bien hacia la derecha o hacia la izquierda. Una medida distinta nos permite verificar eso mismo en imágenes con contraste suficiente, *figura 5.9(e)*, esto podría servir como una medida adicional para verificar que la imagen ha sido procesada correctamente en caso de que la distancia es suficientemente pequeña. Con respecto al criterio de convergencia se puede ver que la convergencia es relativamente rápida,  $J_b$  decrece drásticamente en las primeras iteraciones, iteración dos en las *figuras 5.9(f)* y *5.10(f)*. Aunque este proceso de convergencia se aplica únicamente en imágenes clasificadas como con contraste insuficiente, podemos afirmar nuevamente que el número de iteraciones es relativamente bajo. Esto confirma que hemos avanzado con respecto al número máximo de iteraciones que establecimos en  $t_{max} = 14$  lo cual valida el valor establecido de los parámetros involucrados en el método *FC* propuesto.



## CÁLCULO AUTOMÁTICO DEL UMBRAL DE PORCENTAJE DE VERDE

### 6.1. Introducción

El presente algoritmo es un procedimiento con el cual se establece de forma automática el valor de un umbral de porcentaje de la componente espectral verde para aplicar posteriormente en la segmentación de la imagen. La segmentación de la imagen la realizamos para quedarnos exclusivamente con los píxeles que pertenecen a plantas para su posterior procesamiento. Dependiendo del valor de este umbral, los píxeles cuyas componentes espectrales RGB queden por debajo son eliminados mientras que aquellos píxeles cuyas componentes espectrales queden por encima del umbral permanecen en la imagen. Así pues, la elección del valor del umbral es un paso crítico a la hora de realizar la extracción de colores en la imagen. Un valor del umbral por debajo del valor apropiado haría que la imagen segmentada apareciese con píxeles que no se corresponden con la realidad, es decir, aparecerían como píxeles verdes píxeles que en realidad no lo son. Por el contrario, un valor del umbral por encima del apropiado eliminaría en la segmentación píxeles de nuestro interés, es decir, suprimiría partes verdes de la imagen que deberíamos tener en cuenta en los procesamientos posteriores.

Existen dos razones por las que hemos tenido que desarrollar este algoritmo que establece automáticamente el valor del umbral a aplicar en la segmentación de la imagen. Por un lado, la necesidad de tener que trabajar en tiempo real, nos obliga a que el ajuste de este umbral tenga que ser totalmente automático y no se pueda establecer manualmente un valor para cada imagen. Por otro lado, al trabajar en exteriores, las imágenes tienen una variación en la intensidad de iluminación. Esta variación es completamente aleatoria puesto que depende de las condiciones atmosféricas

(condiciones que varían constantemente) e incluso de los ángulos de orientación de la cámara según el tractor se mueve en una dirección con el sol a favor o en contra, a medio día o al atardecer, etc. Igualmente, la aparición de sombras repentinas en la imagen hace que los valores de la intensidad de los píxeles cambien radicalmente de una imagen con sombras a la siguiente sin ellas. Por ello, un algoritmo que trabajase asumiendo un rango de valores fijo establecido por programa no sería práctico en el caso que nos ocupa. Esto es debido a que un valor es apropiado para unas condiciones de iluminación determinadas pero cuando cambian esas condiciones, dicho valor sería inapropiado porque quedaría por encima o por debajo del óptimo.

Así pues, cada imagen ha de ser tratada de forma independiente respecto del resto y el valor del umbral establecido ha de ser calculado para todas y cada una de las imágenes por separado y en tiempo real. Una vez aplicado el procedimiento, el valor calculado para el umbral será utilizado exclusivamente sobre esa imagen. A continuación se vuelve a aplicar el procedimiento a la siguiente imagen y así sucesivamente.

El procesamiento de las imágenes llevado a cabo para la detección de los porcentajes de malas hierbas utilizado en el programa RHEA (*RHEA 2014*) consta de seis etapas:

1. Adquisición de la imagen.
2. Ajuste del tiempo de exposición para la siguiente imagen.
3. Valoración de la calidad de la imagen.
4. Segmentación de la imagen (extracción de verdes).
5. Detección de líneas de cultivo.
6. Cálculo del porcentaje de malas hierbas.

El algoritmo para la determinación automática del umbral de porcentaje de las componentes espectrales forma parte del *punto cuatro* señalado previamente *como segmentación de la imagen*. Es válido para la extracción de cualquier color. Si se trata de un color puro el que queremos extraer, es decir, con sólo una de las tres componentes espectrales RGB con valor mayor que cero y las otras dos nulas, la segmentación se aplicaría con el valor del umbral aplicado a la componente espectral que es mayor que

cero. En caso de colores multispectrales reales habría que establecer tres valores para tres umbrales diferentes y realizar la segmentación canal por canal, es decir, aplicando un umbral a cada uno de los canales espectrales.

Dado que este algoritmo ha sido expresamente desarrollado dentro del proyecto RHEA (*RHEA 2014*), nuestro interés se centra únicamente en los porcentajes de verde de cada píxel. Por tanto nos centramos exclusivamente en el canal espectral verde G para su aplicación.

Una vez establecido el valor del umbral del porcentaje, mediante el procedimiento que detallamos a continuación, se aplica a la imagen, eliminando todos los píxeles cuyo porcentaje de verde queda por debajo del umbral. El resultado obtenido es una imagen con píxeles cuyo porcentaje de verde es lo suficientemente alto como para poder ser considerados píxeles pertenecientes a objetos verdes, en nuestro caso vegetación. A continuación se procedería con los *puntos cinco y seis* mencionados previamente.

## 6.2. Estado del arte

En la *sección 5.2 “Estado del arte”* se incluye una relación de métodos relativos a la obtención de un umbral con el propósito de segmentar las imágenes para identificar las partes verdes del resto. Dichos métodos caben en esta misma sección en tanto en cuanto el método que se describe a continuación tiene precisamente esa finalidad.

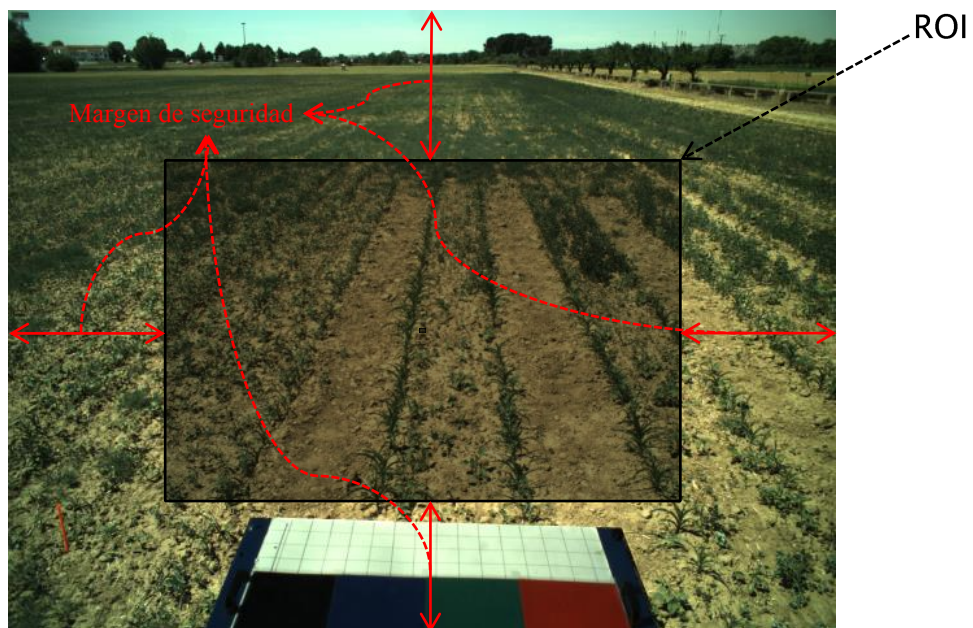
## 6.3. Descripción del método

El método para el cálculo automático del umbral de porcentaje de verde de una imagen consta de cinco pasos que pasamos a describir seguidamente.

1. Una vez adquirida la imagen y clasificada como apta para su procesamiento, se delimita lo que previamente hemos denominado área de interés (ROI) sobre la que trabajar (*figura 6.1*). Existen dos razones por las cuales es necesario establecer dicha área de interés en el interior de la imagen:
  - Debido a la posición de la cámara y a su ángulo de inclinación, la zona central de la imagen es la que coincide con los surcos de cultivo. Por

ello, esta zona sirve de referencia para extraer los datos necesarios para la elección del valor del umbral que aplicamos posteriormente. En caso de no delimitar el área de procesamiento de la imagen, podrían entrar en el cómputo del umbral elementos que falsearían el valor del mismo y la imagen, una vez segmentada, presentaría como elementos verdes objetos que no lo son, por ejemplo, tierra, ramas, cielo, casas, nubes, etc.

- Debido al movimiento del tractor es necesario mantener un “margen de seguridad” alrededor de la zona que se va a procesar. Así pues, a pesar de los movimientos del tractor, siempre tendremos un área fija en la que trabajar que delimita los surcos, sin peligro de que éstos desborden la imagen. Este margen de seguridad es la zona de la imagen que queda fuera de la ROI en la *figura 6.1*.



**Figura 6.1:** Cada imagen recibida para procesar se delimita con un área de interés ROI (Region Of Interest) para realizar el procesamiento en los píxeles incluidos dentro. El resto de la imagen que queda fuera de la ROI no es utilizada durante el procesamiento.

2. Calculamos el porcentaje de verde de cada píxel incluido en la ROI. Para ello, utilizamos las componentes espectrales verde de la imagen RGB, obteniendo una matriz  $P$ , con las mismas dimensiones que la ROI cuyos valores son los porcentajes de verde (normalizados) de cada uno de los píxeles:

$$P_{ij} = \frac{R_{ij}}{R_{ij} + G_{ij} + B_{ij}} \quad (6.1)$$

donde  $P_{ij}$  es el porcentaje de verde (normalizado) del píxel de la fila  $i$  y la columna  $j$  dentro de la ROI.

- De cada columna de la matriz  $P$  se calcula, columna por columna, el valor medio  $m_j$  de los porcentajes:

$$m_j = \frac{1}{f} \sum_{i=1}^f P_{ij} \quad j = 1, \dots, c \quad (6.2)$$

donde  $m_j$  es la media de los porcentajes de la columna  $j$  de la matriz  $P$ ,  $f$  es el número total de filas y  $c$  es el número total de columnas de  $P$ .

- A continuación obtenemos el valor máximo y mínimo de las medias calculadas anteriormente en el *punto 3*:

$$Max = \max \{m_j\};$$

$$Min = \min \{m_j\};$$

- Finalmente establecemos el valor del umbral de porcentaje de verdes,  $U$  para aplicar en segmentación como la media aritmética entre el máximo y el mínimo

$$U = \frac{Max + Min}{2} \quad (6.3)$$

Este valor de umbral  $U$  es aplicado a la imagen de tal forma que los píxeles cuya componente espectral verde esté por debajo del mismo son eliminados de la imagen binarizada y aquellos cuya componente espectral esté por encima o sea igual permanecen en la imagen. El resultado es una imagen binarizada donde todos los píxeles presentes pertenecen a plantas ya sean de cultivo o malas hierbas.

## 6.4. Resultados

El algoritmo ha sido probado en más de 300 imágenes estáticas conteniendo líneas de cultivo de maíz, así como en imágenes variadas conteniendo objetos verdes,

entre otros colores. Igualmente ha sido probado en imágenes de video captadas a razón de dos imágenes por segundo aproximadamente, en los campos de experimentación de La Poveda, Arganda, pertenecientes al Instituto de Ciencias Agrarias del CSIC. Las imágenes de video han sido tomadas en distintos días con diferentes condiciones meteorológicas, sumando un total de más de 4000 imágenes.

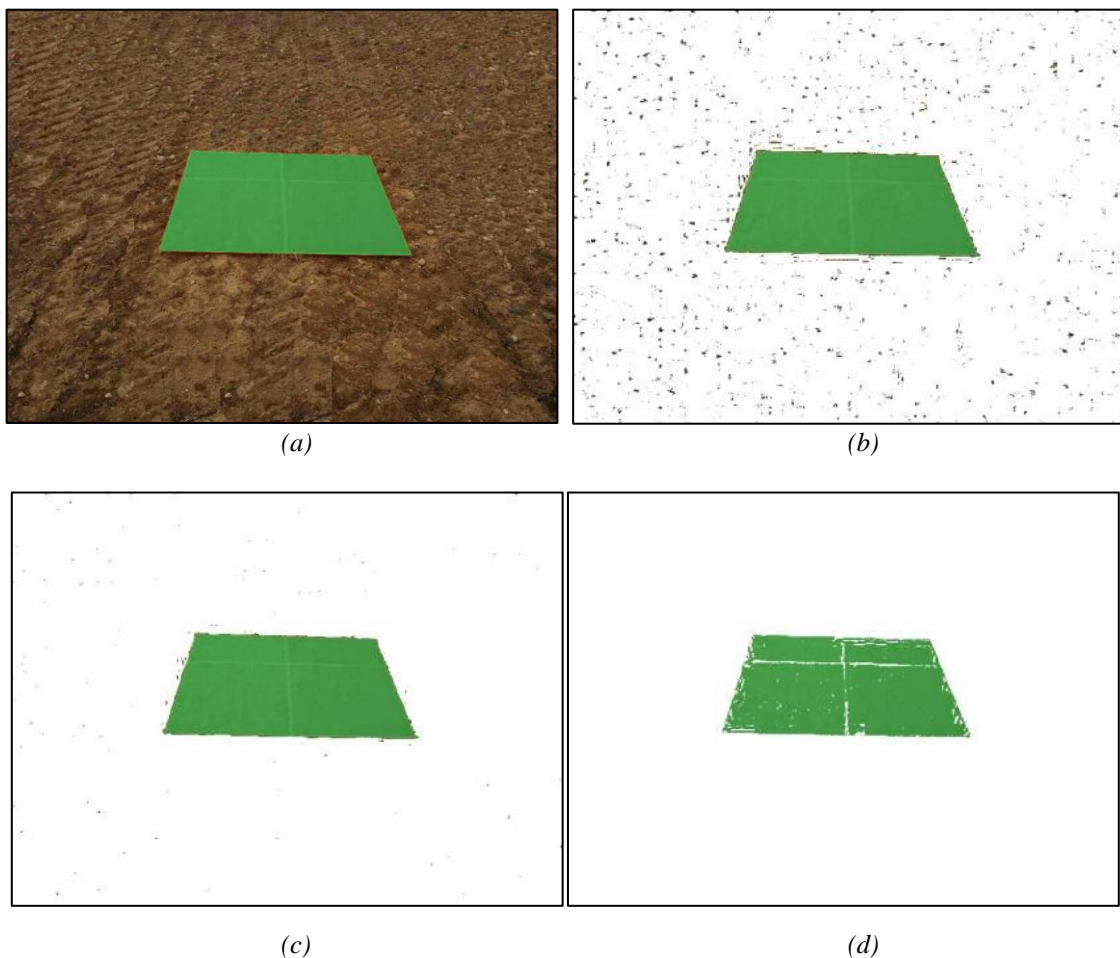
Los resultados de la aplicación del umbral calculado mediante este algoritmo han sido validados mediante inspección visual ya que no existe valor teórico correcto para cuantificar dichos resultados cualitativos.

Se da como valor correcto del umbral aquel cuya aplicación extrae los píxeles verdes apropiadamente y se comprueba visualmente que, si se aumenta o disminuye en dos o tres puntos porcentuales el valor del umbral, la imagen segmentada empeora considerablemente eliminando píxeles verdes o manteniendo píxeles que no son verdes respectivamente.

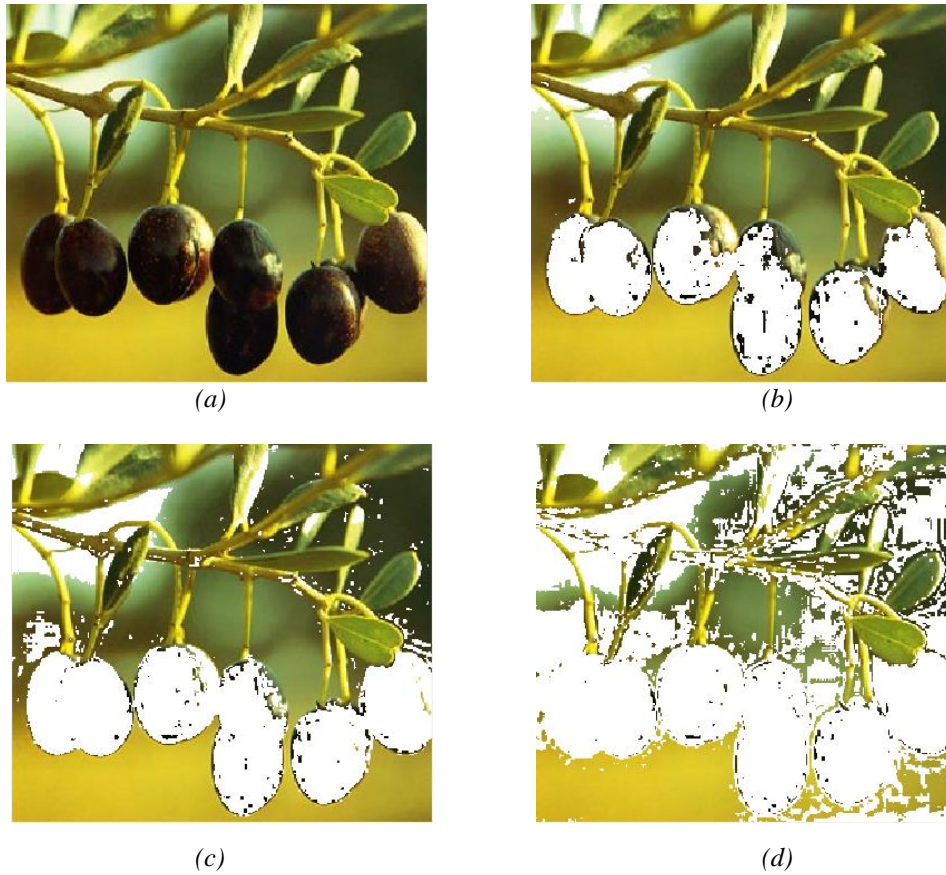
Así pues, la segmentación de un color basada en porcentajes de ese color no tiene un único valor correcto aunque sí es cierto que el rango de valores aceptables (visualmente) es muy pequeño; es decir, si una imagen queda bien segmentada con un valor de umbral de, por ejemplo 36%, es probable que con un 35% o un 37% también quede bien segmentada, pero alejándonos un punto porcentual más o dos puntos, la segmentación comienza a fallar considerablemente. Si nos alejamos entre 4 y 5 puntos porcentuales los resultados de la segmentación empiezan a ser catastróficos ya que en el caso de quedarnos por debajo del umbral apenas eliminaría los píxeles pertenecientes al suelo del terreno y si nos quedásemos por encima eliminaría importantes proporciones de vegetación. Así pues, en general son aceptables entre 1 y 4 puntos porcentuales correlativos en cada segmentación.

A continuación se muestran varios ejemplos de los resultados obtenidos una vez aplicada la segmentación con el valor del umbral calculado de acuerdo con el método propuesto. Las imágenes etiquetadas con *(a)* son las imágenes originales. Las etiquetadas con *(c)* son las imágenes una vez segmentadas con el valor del umbral calculado por nuestro algoritmo. Las imágenes con *(b)* son imágenes segmentadas con un valor del umbral 2 ó 3 puntos porcentuales por debajo del calculado por nuestro algoritmo. Finalmente las imágenes etiquetadas con *(d)* son el resultado de segmentar la

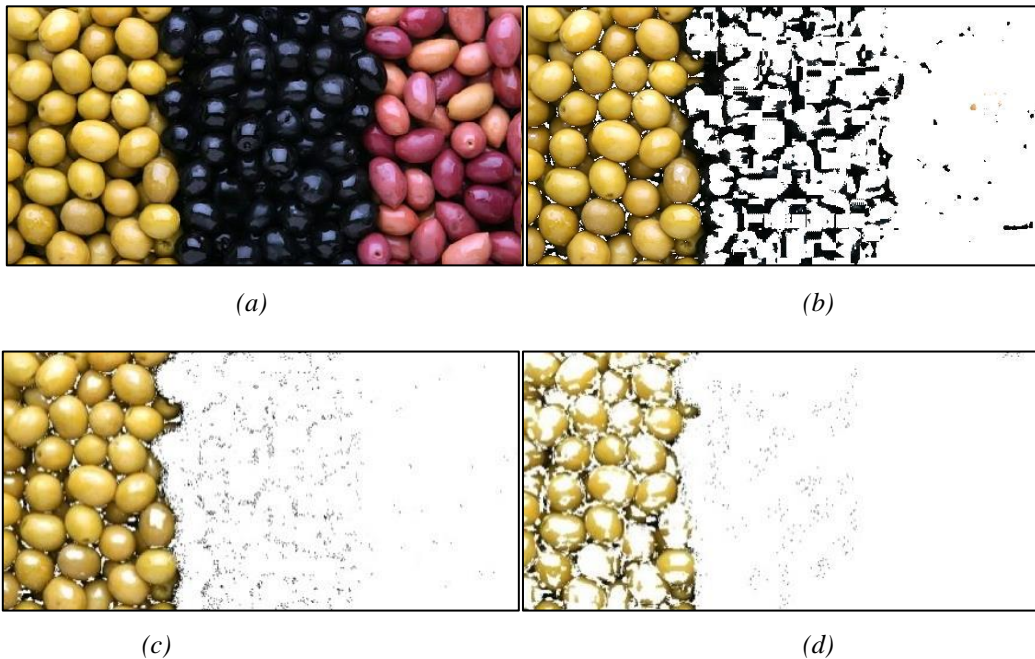
imagen original con un umbral 2 ó 3 puntos porcentuales por encima del valor calculado por nuestro algoritmo. En todos los casos puede observarse cómo el valor calculado por el método propuesto se encuentra dentro de los valores aceptables para una segmentación correcta mientras que si nos alejamos 2 ó 3 puntos porcentuales por debajo o por encima de ese valor comienzan a permanecer en la imagen píxeles que no son verdes o a desaparecer píxeles verdes respectivamente.



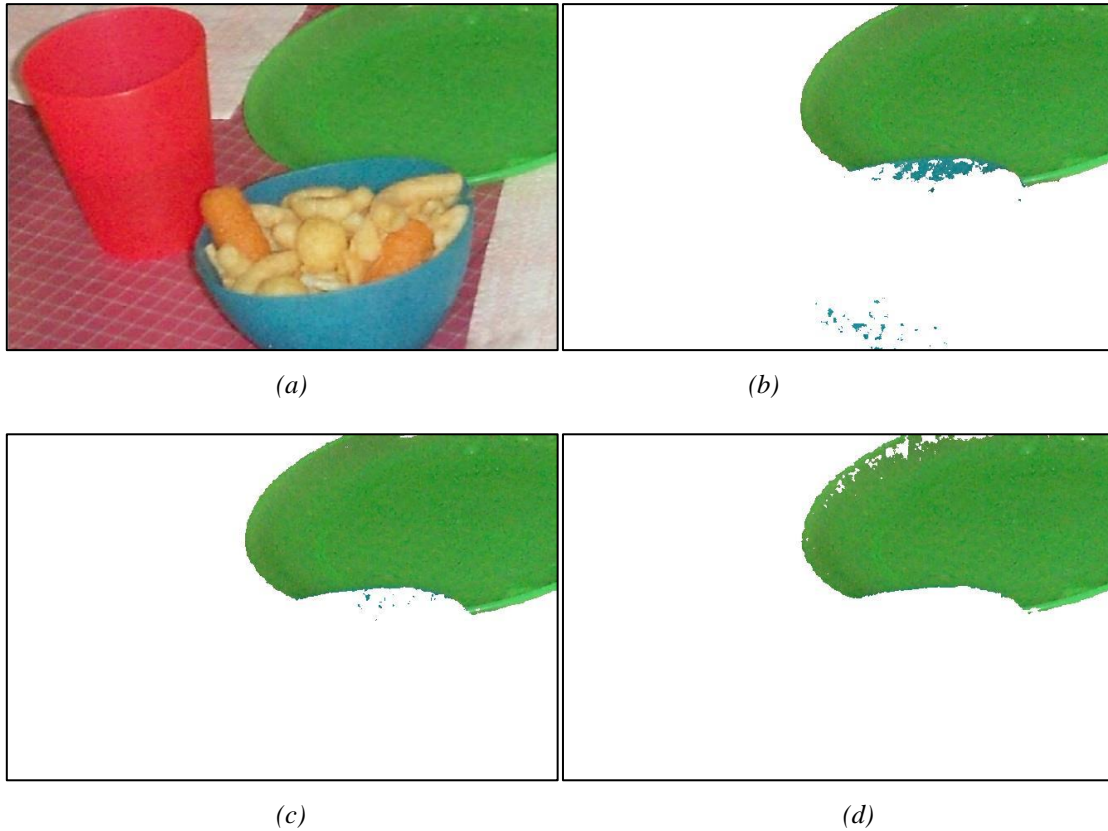
**Figura 6.2:** (a) Tabla de un metro cuadrado de color verde colocada sobre tierra sin sembrar. (b), (c) y (d) muestran la imagen segmentada aplicando un umbral de valor 36%, 38% y 40% respectivamente. (c) Muestra el valor dado por el algoritmo. Obsérvese que por debajo de ese valor, es decir con un 36%, empiezan a aparecer como píxeles verdes píxeles que no lo son (b). Igualmente incrementando hasta 40 el umbral, píxeles verdes empiezan a desaparecer de la imagen segmentada (d).



**Figura 6.3:** (a) Imagen original de aceitunas negras sobre fondo verde; (b), (c) y (d) muestran la imagen segmentada aplicando un umbral de valores 36%, 39% y 42% respectivamente, con el 39% como mejor valor. Obsérvese que por debajo de ese valor, es decir con un 36%, empiezan a aparecer como píxeles verdes puntos que no lo son (b). Igualmente incrementando hasta 42% el umbral, píxeles verdes empiezan a desaparecer de la imagen segmentada (d).



**Figura 6.4:** (a) Imagen original de aceitunas verdes, negras y moradas. (b), (c) y (d) muestran la imagen segmentada aplicando un umbral de valor 36%, 38% y 40% respectivamente. (c) Muestra es el valor dado por el algoritmo. Obsérvese que por debajo de ese valor, es decir con un 36%, empiezan a aparecer como píxeles verdes puntos que no lo son (b). Igualmente incrementando hasta 40% el umbral, píxeles verdes empiezan a desaparecer de la imagen segmentada (d).



**Figura 6.5:** (a) Imagen original de objetos de colores. (b), (c) y (d) muestran la imagen segmentada aplicando un umbral de valor 42%, 45% y 48% respectivamente. (c) Muestra es el valor dado por el algoritmo. Obsérvese que por debajo de ese valor, es decir con un 42%, empiezan a aparecer como píxeles verdes puntos que no lo son (b). Igualmente incrementando hasta 48% el umbral, píxeles verdes empiezan a desaparecer de la imagen segmentada (d).



## CONCLUSIONES

A lo largo del presente trabajo se han expuesto diferentes métodos con diferentes propósitos orientados a dar soluciones a los problemas planteados bajo los proyectos RHEA (*RHEA 2014*) y AGL (2008, 2011). Las conclusiones más relevantes se presentan bajo diferentes perspectivas, a saber: *a)* detección de líneas de cultivo en distintos campos; *b)* cálculo de coordenadas de objetos en imágenes; *c)* ajustes del sistema de visión para detección de líneas e identificación de malas hierbas con la máxima precisión posible; *d)* desarrollo de un sistema experto para identificación de verdes con fines de segmentación de la imagen; *e)* cálculo automático del umbral en la imagen del porcentaje de verdes. Estos métodos conforman un conjunto de procedimientos basados en técnicas propias del tratamiento de imágenes y la visión por computador aplicadas en agricultura, o si se prefiere, en lo que se conoce como agricultura de precisión, tanto en lo relativo a tratamientos específicos con el fin de proporcionar la información suficiente para eliminar tanto malas hierbas en campos de cultivo como ayuda al guiado de vehículos autónomos.

A pesar de los resultados satisfactorios obtenidos, como cualquier técnica que se precie es susceptible de mejora en el futuro. Aspecto éste que se aborda en el siguiente capítulo. También en dicho capítulo se considera la posibilidad de la aplicación de los métodos desarrollados a otras tareas en el ámbito agrícola e incluso alguno de los métodos propuestos podría traspasar dicho ámbito debido a su propia naturaleza, tal es el caso de la detección de líneas de cultivo como ayuda al guiado.

A continuación se detallan las conclusiones más relevantes de cada uno de los algoritmos desarrollados descritos en la presente memoria.

## 7.1. Algoritmo para la detección automática de líneas de cultivo

Hemos desarrollado un nuevo método para la detección de líneas de cultivo que mejora considerablemente los resultados que se obtienen aplicando algunos métodos clásicos, como es el método basado en la transformada de Hough. La mejora se ha conseguido en términos de tiempo de computación, robustez y fiabilidad. Tanto las diferencias en la tipología de imágenes, esto es, las variaciones provocadas por los movimientos de la cámara en el tractor así como la presencia masiva de malas hierbas son soportados suficientemente por el procedimiento propuesto. El método consta de dos pasos: extracción de verdes y detección de líneas de cultivo. La extracción de verdes es básicamente un método de segmentación, que utiliza las tres componentes espectrales en el modelo de color RGB de cada píxel. Está basado en técnicas de aprendizaje que nos proporciona un umbral único válido para toda la imagen.

La detección de líneas de cultivo se basa en un método nuevo, rápido y robusto desarrollado exclusivamente para imágenes con proyección de perspectiva. Es válido tanto para imágenes de alta como de baja resolución. Detecta cualquier número de líneas de cultivo con cualquier pendiente de línea, siempre que dichas líneas converjan según dicha proyección, lo cual ocurre teóricamente en el denominado punto de fuga.

## 7.2. Cálculo de las *coordenadas x-y* de objetos en imágenes

Se ha desarrollado un método que no sólo determina la distancia lineal de un objeto respecto al observador sino que calcula las *coordenadas x-y* del objeto respecto a un sistema de coordenadas ubicado por el observador. En nuestro, caso el sistema de coordenadas se ha colocado coincidente con la cámara y los ejes de tal forma que el *eje-X* es paralelo a las filas de la imagen y coincidente con la primera fila de píxeles de ésta; el *eje-Y* es paralelo a las columnas de píxeles y coincidente con la columna ubicada en la mitad de la imagen. Este método es únicamente válido para objetos situados sobre superficies planas como pueden ser suelos horizontales sin pendientes, campos lisos, superficies acuáticas, etc.

Está diseñado para utilizar una única cámara de la cual no es necesario conocer ni sus parámetros intrínsecos ni extrínsecos. La información proporcionada por dichos

parámetros es sustituida por una calibración realizada “in situ” con la cámara colocada en una posición fija. Se toman distintas imágenes cuyas *coordenadas x-y* se obtienen mediante GPS. Utilizando esta información el método proporciona unas funciones de propósito general  $F_y(r)$  y  $F_x(c,r)$ , válidas para cualquier otro punto de la imagen, de suerte que mientras no se mueva la cámara, sirven para obtener la posición del objeto (*coordenadas x-y*) en función de la fila y la columna en la que aparece dicho objeto sobre la imagen. El número de muestras necesario para poder obtener unas funciones  $F_y(r)$  y  $F_x(c,r)$  fiables depende de la precisión en las *coordenadas x-y*, la resolución de las imágenes y del campo de visión que queramos cubrir con dichas funciones.

Este método completa los existentes actualmente, que se centran en el cálculo de la distancia lineal del objeto al observador ya que, además de calcular dicha distancia, proporciona las *coordenadas x-y* del objeto respecto a un origen de coordenadas establecido por el observador.

### **7.3. Ajustes de la cámara para detección de líneas de cultivo y malas hierbas con la máxima precisión**

El presente método fue la adquisición de imágenes con la mayor precisión posible para la detección de líneas de cultivo y de malas hierbas, una vez que los parámetros intrínsecos de la cámara estaban definidos, es decir, una vez que se había elegido qué cámara usar. Este método fue inspirado para resolver específicamente la problemática en el proyecto *RHEA (2014)*.

Hemos observado dos fuentes de imprecisiones. La primera es debida a una incorrecta elección de los valores de los parámetros extrínsecos una vez que los parámetros intrínsecos han sido fijados de acuerdo con los requisitos propios de la aplicación. Por ejemplo que la ROI debe estar situada a cierta distancia del tractor de tal forma que haya tiempo suficiente para procesar la imagen en tiempo real con el fin de proceder al tratamiento adecuado. Igualmente, un ángulo de inclinación podría hacer desaparecer alguna de las líneas de cultivo de nuestro interés. Otra cuestión a considerar es el hecho de que a pesar de que existe más resolución y por tanto más precisión colocando la ROI en la parte superior o inferior de la imagen, no es posible ubicarla donde sería deseable hacerlo debido a los movimientos constantes del tractor cuando está en marcha. Esto es así ya que debemos dejar un margen de seguridad de tal forma

que la ROI siempre quede dentro de la imagen incluso con los movimientos provocados por la orografía del terreno. En cada proyecto los parámetros intrínsecos y extrínsecos tendrán que combinarse de tal forma que se llegue a un compromiso entre todos ellos para lograr la mayor precisión posible. La segunda fuente de imprecisiones estudiadas han sido las derivadas de una iluminación no controlada en entornos de exterior, la cual puede excitar en exceso o insuficientemente el sensor provocando saturación o todo lo contrario. Por último, conviene señalar el conocido como efecto *vignetting* provocado al atravesar los rayos de luz las distintas capas de los filtros UV-IR.

En todos los casos mencionados hemos propuesto soluciones para corregir los efectos adversos de una iluminación no controlada con el objetivo de lograr la máxima precisión posible. Concluimos pues, que para procesamiento de imágenes en escenarios de exterior, la iluminación que accede al sensor debe estar permanentemente controlada en tiempo real, es decir, según son captadas las imágenes y que los umbrales y tiempos de exposición aplicados en cada caso no son valores únicos sino que éstos dependerán de la sensibilidad que requiera la aplicación.

## **7.4. Sistema experto para identificación de verdes en imágenes agrícolas**

Hemos propuesto un nuevo sistema experto automático para la identificación del verde en imágenes agrícolas en campos de maíz y cereal. El diseño abierto de su arquitectura permite que pueda ser aplicado a cualquier tipo de cultivo desde el momento en que comiencen a crecer las plantas del cultivo en cuestión. Una importante ventaja de este sistema experto es que también es aplicable a otros entornos donde el objetivo sea la detección de zonas verdes, como por ejemplo, imágenes aéreas de áreas urbanas en las que el cálculo del porcentaje de las zonas verdes, en comparación con las zonas urbanizadas, es un dato de gran interés a la hora de catalogar esa zona desde un punto de vista medioambiental. Otro ejemplo en el que el sistema experto podría ser igualmente utilizado es en las imágenes aéreas de bosques o zonas arboladas. Aquí el interés está nuevamente en una cuantificación aproximada de masa forestal y densidades.

El sistema experto propuesto está basado en dos módulos, donde el primero clasifica la calidad de la imagen mediante un análisis del histograma. De

acuerdo con el resultado de esta clasificación la imagen es procesada o bien aplicando una combinación clásica de métodos basados en la extracción de verdes o bien mediante un agrupamiento difuso para establecer un umbral dinámico para cada imagen.

La estrategia propuesta ha sido probada con una multitud de imágenes obtenidas con distintas cámaras, con el objetivo de desarrollar aplicaciones automáticas en imágenes agrícolas para tratamientos específicos y capturadas bajo condiciones medioambientales muy variadas.

El sistema experto ha sido desarrollado con una arquitectura abierta de tal forma que en un futuro sea posible sustituir o añadir nuevos módulos, siendo de particular interés para el caso que nos compete el de métodos automáticos de establecimiento de umbrales (Avci y Avci 2009) así como mejoras en la regla utilizada para la toma de decisión de la ecuación (5.3) mediante la aplicación de un árbol de decisión con estrategia inductiva de aprendizaje como por ejemplo ID3, CART o C4 entre otros (Duda y col. 2000). Para el desarrollo de este sistema experto hemos aplicado la transformada discreta de wavelets a los tres canales espectrales R, G y B. El proceso de submuestreo podría aplicarse sólo a la imagen de intensidad tras una transformación de la imagen desde el espacio de colores RGB al espacio HSI (Hue, Saturation Intensity) entre otros.

Además, una vez que las partes verdes han sido identificadas, las partes restantes (principalmente las pertenecientes al suelo) podrían ser analizadas para identificar categorías ecológicas relevantes (Luscier y col. 2006). De esta forma, el SE propuesto podría ser extendido para analizar los materiales de la tierra con la retroalimentación de muestras pertenecientes a esta categoría

## **7.5. Algoritmo para el cálculo automático del umbral de porcentaje de verde**

Se ha propuesto un algoritmo para la segmentación de verdes en tiempo real cuyos resultados quedan dentro del rango de valores aceptables para su posterior procesamiento. Dado que no existe un único valor correcto para este umbral de corte, los resultados han de ser validados mediante inspección visual y opinión subjetiva dada por un experto. Sin embargo, no es difícil determinar la validez de dicho valor dado que

el rango de valores posibles del umbral es realmente estrecho. Por lo general, son aceptables de tres a cinco valores de porcentaje. Fuera de ese rango de valores, la segmentación es excesiva o insuficiente dependiendo de que el valor esté por encima o por debajo del rango respectivamente. Con este algoritmo, los resultados quedan dentro del rango de valores válidos para la segmentación de verdes en la totalidad de las imágenes analizadas (más de 300 imágenes estáticas y 4000 de vídeo).

## TRABAJOS FUTUROS

Los métodos anteriores se han desarrollado bajo la perspectiva de los proyectos *RHEA (2014)* y *AGL (2008, 2011)* donde el objetivo fundamental ha sido la aplicación de tratamientos selectivos en campos de cultivo durante la época de crecimiento de los mismos, donde tanto los cultivos como las malas hierbas se encuentran en un estado vegetativo verde. En este sentido se plantean dos tipos de líneas de trabajo futuro que se describen seguidamente.

### 8.1. Mejora de los métodos y algoritmos propuestos

Debido a las limitaciones en el tiempo de cómputo provocadas por la necesidad de trabajar en tiempo real, el algoritmo para la detección de líneas de cultivo descrito en el capítulo dos está desarrollado para detectar exclusivamente las líneas de cultivo que nacen en la base inferior de la imagen. Sin embargo, como trabajo futuro sería útil para otras aplicaciones implementar el algoritmo de tal forma que también detectase las líneas de cultivo que nacen en el lateral izquierdo y derecho de la imagen.

En relación al método para detección de objetos en imágenes, sería conveniente establecer un criterio para definir cuál es el píxel cuya fila y columna se van a utilizar para determinar la ubicación del objeto. Esto viene motivado porque en el método propuesto simplemente se explica que tiene que ser un píxel en contacto con la superficie de la escena. Igualmente para poder determinar automáticamente el píxel que va a representar al objeto es necesario realizar una limpieza en la imagen de todos aquellos píxeles que no pertenezcan al objeto. Esto lleva a desarrollar un algoritmo específico para cada aplicación según sea la superficie sobre la que se sitúe el objeto.

En cuanto al método relacionado con los ajustes de la cámara para lograr la máxima precisión posible, se propone, el establecimiento de una base de datos similar a

la base de datos de imágenes estereoscópicas de Middlebury (2014) de *Scharstein y Pal (2007)* pero de imágenes de exterior donde se conozca la imagen de verdad en cada caso o “ground-truth” de los píxeles verdes presentes en cada imagen y sus componentes espectrales de acuerdo con un único patrón. De esta forma tanto éste como cualquier otro algoritmo de segmentación podrían ser validados de una forma cuantitativa.

En relación al sistema experto (SE) desarrollado, podría adaptarse con relativa facilidad para la clasificación de otros elementos, distintos a las plantas verdes, presentes en las imágenes para su clasificación e identificación. En este sentido podría utilizarse en análisis de suelo o de las plantas en fases de crecimiento donde el verde deja de ser el color predominante.

En relación al método de detección automática del umbral en el porcentaje de verde, el método utiliza técnicas estadísticas para calcular el promedio, concretamente medias aritméticas. Sin embargo, también conviene explorar otro tipo de estadísticos, dado que existen indicios suficientes respecto de su validez, como son medias geométricas o varianzas.

Sería también conveniente, tal y como se ha señalado previamente, crear una base de imágenes de “ground-truth” para poder comparar resultados de forma cuantitativa y poder así depurar y mejorar los métodos utilizados.

## **8.2. Adaptación de los métodos para su utilización en otras tareas**

Los métodos descritos en la presente memoria se han diseñado específicamente para su utilización en aplicaciones agrícolas donde el tema central es la identificación de verdes con distintos fines. Es evidente, que en el propio ámbito agrícola existen diversas tareas en las que por su naturaleza no existen plantas verdes para su identificación, bien porque se están realizando tareas preparatorias del terreno para proceder a la siembra de cultivos o bien porque éstos se encuentren en fase de cosecha o ya cosechados. En ambos casos, también existen procesos de mecanización y utilización de vehículos autónomos. Sería interesante la adaptación de los métodos de procesamiento de imágenes propuestos, para su aplicación en tareas automáticas sin presencia de plantas verdes como base de la aplicación.

Incluso trascendiendo el ámbito agrícola, los métodos propuestos podrían adaptarse perfectamente con fines de identificación de determinados tipos de elementos con tonalidades diferentes al verde o para la detección de otros tipos de líneas diferentes a las de cultivo, lo cual es útil por ejemplo para el guiado autónomo de vehículos en entornos no agrícolas.

La discriminación de distintos tipos de plantas verdes mediante la utilización de técnicas basadas en la obtención de un doble umbral para diferenciar diversas tonalidades de verdes constituye también una tarea de interés, que ya ha sido abordada en el grupo de investigación donde se ha desarrollado el presente trabajo, si bien requiere todavía ciertas mejoras.

Finalmente, dentro del ámbito agrícola, la distinción de plantas por la forma, mediante la segmentación previa de los verdes, representa otra tarea de futuro. Concretamente, la identificación de malas hierbas de los tipos monocotiledóneas o dicotiledóneas constituye otro de los retos de futuro en el ámbito de los métodos propuestos en este trabajo.

UNIVERSIDAD COMPLUTENSE DE MADRID  
Facultad de Ciencias Físicas



Departamento de Arquitectura de Computadores y Automática

## **PhD Thesis**

# ***Computer Vision Methods for objects and textures localization and identification in outdoors***

D. Juan Sebastián Romeo Granados

Supervised by:

Dr. D. Gonzalo Pajares Martinsanz  
Dr. D. Jesús Manuel de la Cruz García

Madrid (Spain), 2014

# Acknowledgments

The research leading to these results has been funded by the European Union's Seventh Framework Programme [FP7/2007–2013] under Grant Agreement no. 245986 in the Theme NMP-2009-3.4-1 (Automation and robotics for sustainable crop and forestry management). The author wish also to acknowledge to the Project AGL-2008-04670-C03-02, supported by the Ministerio de Ciencia e Innovación of Spain within the Plan Nacional de I+D+i.

There are many people that have made this work possible. It would not be possible to name them all without missing one of them. However I would like to express my sincere gratitude especially to my thesis directors Gonzalo Pajares and Jesus Manuel de la Cruz for their great support and help provided during this years.

Of course I cannot forget my family: Javier, Alejandro, Juan and Esther without their support and patience it would have not been possible to do this work.

# INDEX

## CHAPTER 1

INTRODUCTION .....	133
1.1. Problem statement .....	133
1.2. Motivation .....	134
1.3. Objectives .....	136
1.4. Methodology.....	137
1.5. Contributions of the research.....	137
1.5.1. Published articles in indexed journals .....	137
1.5.2. Conferences and Workshops .....	139
1.5.3. Lectures and Seminars.....	139
1.5.4. Radio broadcasts.....	140
1.6. English summary organization .....	140

## CHAPTER 2

CROP ROW DETECTION IN MAIZE FIELDS INSPIRED ON THE HUMAN VISUAL PERCEPTION .....	141
2.1. Problem statement .....	141
2.2. Revision of methods .....	142
2.2.1. Methods Based on the Exploration of Horizontal Strips.....	142
2.2.2. Methods Based on the Hough Transformation.....	143
2.2.3. Vanishing Point Based .....	144
2.2.5. Stereo Based Approaches.....	145
2.2.6. Methods Based on Blob Analysis.....	146
2.2.7. Methods Based on the Accumulation of Green Plants.....	146
2.2.8. Methods Based on Frequency Analysis.....	146
2.3. Motivational research and design of the proposed strategy .....	147
2.4. Materials and Methods .....	148
2.4.1. Images.....	148
2.4.2. Image Segmentation: Green Plants Identification .....	149
2.4.3. Crop Row Identification .....	152
2.5. Results.....	155
2.6. Conclusions.....	159

## CHAPTER 3

AUTONOMOUS ROBOTS POSITIONING IN WATER SURFACE FROM A SINGLE IMAGE.....	161
3.1. Introduction .....	161

3.2 State of the art.....	162
3.2.1. Stereopsis.....	163
3.2.2. Motion .....	164
3.2.3. Shading.....	164
3.2.4. Texture.....	164
3.2.5. Focus/Defocus .....	165
3.2.6. Depth estimation with a single camera.....	165
3.3. Method description.....	167
3.4. Experimental results .....	173
3.5. Conclusions. ....	176
<b>CHAPTER 4</b>	
<b>CAMERA SENSOR ARRANGEMENTS FOR CROP AND WEED DETECTION ACCURACY IN AGRONOMIC IMAGES.....</b>	
	177
4.1. Introduction .....	177
4.2. Materials .....	179
4.3. Methods .....	181
4.3.1. Accuracy from the point of view of extrinsic parameters .....	181
4.3.2. Accuracy from the point of view of illumination factors .....	184
4.3.2.1. Exposure Time.....	185
4.3.2.2. Vignetting correction.....	186
4.4. Results and discussion.....	187
4.4.1. Analysis of extrinsic parameters .....	188
4.4.2. Analysis of illumination factors .....	193
4.5. Conclusions .....	197
<b>CHAPTER 5</b>	
<b>A NEW EXPERT SYSTEM FOR GREENNESS IDENTIFICATION IN AGRICULTURAL IMAGES.....</b>	
	199
5.1. Introduction .....	199
5.2. State of the art.....	200
5.3. Motivational research of the proposed strategy.....	202
5.4. Expert system design.....	203
5.4.1. Decision making.....	204
5.4.2. Greenness identification .....	205
5.4.2.1. Combination of methods .....	206
5.4.2.2. Preprocessing and Fuzzy Clustering .....	207
5.5. Results .....	209
5.5.1. Decision making module: identification of parameters.....	210
5.5.2. Greenness Identification Module.....	214

5.6. Conclusions .....	223
CHAPTER 6	
AUTOMATIC GREENNESS THRESHOLD DETECTION.....	225
6.1. Introduction .....	225
6.2. Description of the method .....	226
6.3. Results .....	228
6.4. Conclusions .....	231
REFERENCES.....	233



## INTRODUCTION

### 1.1. Problem statement

In computer vision there exist two kinds of environments for images: structured and non-structured. The first ones are mainly located in the industrial field where most of the external variables are perfectly controlled. Control quality processes are an example of this type of environments. In these processes, manufactured products on a conveyor belt are supervised by a computer vision system to approve their quality. Extrinsic variables are perfectly controlled and image processing algorithms must only beware of intrinsic variables. On the other hand, in non-structured environments external variables present a random behavior and there is no control over them. Natural environments belong to this group. Images acquired from natural environments are subject to atmospherical conditions and variables such as illumination, brightness, infrared and ultraviolet radiation among others are out of any human control.

Agricultural and aquatic applications are two clear examples of non-structured applications of computer vision where environmental conditions are highly unpredictable and changeable. Therefore, algorithms for images processing must be designed and developed with robustness enough to cope with this fact. Generally, problems must be considered from a completely point from depending whether we are in a structured environment or in a non-structured environment.

The present work deals with agricultural and aquatic environments to develop different algorithms that are widely explained in the next chapters.

In the agricultural environments, a tractor equipped with a vision system on its top moves along maize crops. Our work has consisted on developing different

algorithms to detect, locate and quantify weeds for a following control treatment. In the aquatic environments we have developed an algorithm to calculate the real coordinates of a USV (Unmanned Surface Vehicle) from an image. In both cases we have to deal with the handicap of the environmental conditions where sunny and cloudy days gave us completely different images of the same scene. In addition to it, in both environments algorithms are to be run on real time which forced us to adjust them to the limit. Finally, and again in both cases we are dealing with non-static scenes which gave us an extra difficulty because of the frame differences between one image and the next one.

## 1.2. Motivation

The present work has its origin in the research projects that are mentioned in the following section. All of them are in the research frame of the ISCAR group (2006) which is officially approved by the Complutense University of Madrid since 2006.

1. The research starts within the frame of the National Plan of R+D 2008 for 2009/2011 granted with the reference number AGL2008-04670-C03-02/AGR and titled: “*PERcepción vIsual, toma de deCIsión y Actuación en sistemas artificiales complejos: control selectivo de malas hierbas en cultivos agrícolas, PERICIA*” (Visual perception and actuation on complex artificial systems: weeds selective control in agricultural crops). In this project, with a work contract associated, stereoscopic techniques for weeds detection in maize and cereal fields were developed for a later herbicide application. The main target of this project was to minimize the adverse effects caused by the use of chemical products as well as to reduce treatment costs. This project was extended in the new project titled: “*Sistema de inspección terrestre en vehículos autónomos y su aplicación efectiva a la detección de malas hierbas y su control localizado* (AGL2011-30442-C02-02)” (ground inspection systems in autonomous vehicles and its affective application for weeds detection and its control). This project was also inside the National Plan of R+D+i for 2011/2014. It is coordinated by the Robotic and Automatic Center (CAR), a mixed center with the Universidad Politécnica de Madrid, as well as the Sustainable Agricultural Institute from CSIC. It has the same target as its predecessor with the only difference that it is focused in techniques for the spatial structures location rather than identification.

2. R+D+i National Plan Project titled: “*Sistema de Vigilancia, Búsqueda y Rescate en el Mar mediante colaboración de Vehículos Autónomos Marinos y Aéreos (DPI2009-14552-C02-01)*” (Surveillance search and rescue system in the sea by autonomous marine and aerial vehicles). As its title says, the target of this Project is to develop marine and aerial vehicles with the aim of locating survivors on the sea. In this project, a constant knowledge of the position between the vehicles implied is needed. This is achieved by a computer vision system which is able to give spatial coordinates of the ships from an established coordinate origin.
  
3. European Community financed Project within the seventh frame (FP7) 2010-2014 titled: *Robot fleets for High Affective Agriculture and Forestry Management (RHEA)*, included into the THEME NMP-2009-3.4-1, (*RHEA 2014*). This project is developed under a consortium made up of nineteen research centers and private companies from eight different European countries. The RHEA project (*RHEA 2014*) is devoted to change the traditional way of proceeding in agriculture and forestry for weed removal and pest management by putting together a fleet of small, safe, reconfigurable, heterogeneous and complementary robots to minimize chemical products, energy and operation time, to maximize the quality of products and the safety and to guarantee the application of the procedures to the entire operation field. One of these robots is shown on *figure 1.1.*, where a tractor is equipped with a vision system for weed detection, location and quantification as well as for tractor guidance.

As we have seen in the previous section, the projects that have motivated this work are being developed in non-structured environments. For this reason, the motivation of this work has been to develop algorithms robust enough to cope with the different hitches derived from these environments.

From a technological point of view, the motivation was to automatize different tasks aimed to work with the maximum efficiency in real time.



*Figure 1.1: Robotized tractor equipped with a vision system on the upper front arch.*

### 1.3. Objectives

In view of the considerations expressed in the previous section, the objectives for image segmentation and objects location are as follows:

1. To develop segmentation techniques in non-structured environments with the aim of identify relevant structures on such environments.
2. To develop methods of identifying relevant elements in the images. In agricultural applications the relevant elements are the crop lines and weeds. In water surface applications the relevant structures are ships navigating on it.
3. To identify the new lines of research to improve the proposed strategies as well as to establish new methods to be used in non-structured environments.
4. To disseminate the results of this research via publications, conferences, congress or any other proper medium.

The fact that all the algorithms have been developed for working on real time has an extra difficulty due to two main reasons:

- a) The lack of computing time forces algorithms to be extremely optimized without losing their robustness and efficiency.
- b) Algorithms for non-structured environments need a constant adjustment of certain parameters according to the changing conditions. These changes cannot

be done manually since they are working on real time. Therefore complementary algorithms for adjusting these parameters had also been developed.

## **1.4. Methodology**

The results presented in this thesis have been achieved following the scientific method directives. In a first stage, an exhaustive bibliographic search was done from the different available sources. Consorcio Madroño network was mainly used for this purpose. Once the information was compiled, the different methods and approaches for similar statements were analyzed to establish the state of the art in this field. The steps were as follows:

1. Problem statement: necessities and motivation.
2. Objectives statement.
3. Definition of methods
4. Obtaining experimental data.
5. Classification, analysis and interpretation of data.
6. General conclusions.
7. Description of future lines of research.

Experimental data have been acquired using the proposed method. This includes the creation of a wide data base, the visualization and analysis of the results and the establishment of the conclusions.

## **1.5. Contributions of the research**

The main contribution of this work is the development of some segmentation techniques for non-structured environments and its following relevant structures identification. We summarize here these works; some of them have been disseminated in indexed international journals.

### **1.5.1. Published articles in indexed journals**

1. *Romeo et al. (2012a)*, this paper proposes a new method, oriented to image real-time processing, for identifying crop rows in maize fields in the images. The vision system is designed to be installed onboard a mobile agricultural vehicle, that is, submitted to gyros, vibrations, and undesired movements. The images are captured

under image perspective, being affected by the above undesired effects. The image processing consists of two main processes: image segmentation and crop row detection. The first one applies a threshold to separate green plants or pixels (crops and weeds) from the rest (soil, stones, and others). It is based on a fuzzy clustering process, which allows obtaining the threshold to be applied during the normal operation process. The crop row detection applies a method based on image perspective projection that searches for maximum accumulation of segmented green pixels along straight alignments. They determine the expected crop lines in the images. The method is robust enough to work under the above-mentioned undesired effects. It is favorably compared against the well-tested Hough transformation for line detection.

2. *Romeo et al. (2013a)*, in Precision Agriculture, images coming from camera-based sensors are commonly used for weed identification and crop line detection, either to apply specific treatments or for vehicle guidance purposes. Accuracy of identification and detection is an important issue to be addressed in image processing. There are two main types of parameters affecting the accuracy of the images, namely: (a) extrinsic, related to the sensor's positioning in the tractor; (b) intrinsic, related to the sensor specifications, such as CCD (Charged Coupled Device) resolution, focal length or iris aperture, among others. Moreover, in agricultural applications, the uncontrolled illumination, existing in outdoor environments, is also an important factor affecting the image accuracy. This paper is exclusively focused on two main issues, always with the goal to achieve the highest image accuracy in Precision Agriculture applications, making the following two main contributions: (a) camera sensor arrangement, to adjust extrinsic parameters and (b) design of strategies for controlling the adverse illumination effects.
3. *Romeo et al. (2013b)*, it is well-known that one important issue emerging strongly in agriculture is related with the automation of tasks, where camera-based sensors play an important role. They provide images that must be conveniently processed. The most relevant image processing procedures require the identification of green plants, in our experiments they comes from barley and maize fields including weeds, so that some type of action can be carried out, including site-specific treatments with chemical products or mechanical manipulations. The images come from outdoor

environments, which are affected for a high variability of illumination conditions because of sunny or cloudy days or both with high rate of changes. Several indices have been proposed in the literature for greenness identification, but under adverse environmental conditions most of them fail or do not work properly. This is true even for camera devices with auto-image white balance. This paper proposes a new automatic and robust Expert System (ES) for greenness identification. It consists of two main modules: (1) decision making, based on image histogram analysis and (2) greenness identification, where two different strategies are proposed, the first based on classical greenness identification methods and the second inspired on the Fuzzy Clustering approach. The Expert System design as a whole makes a contribution, but the Fuzzy Clustering strategy makes the main finding of this paper. The system is tested for different images captured with several camera devices.

4. The following is a list of relevant contributions as co-author in journals related to the research: *Guerrero et al. (2012a, 2013)* and *Montalvo et al. (2012a, 2013a)*.

### **1.5.2. Conferences and Workshops**

5. *Romeo et al. (2012b)*, a new method for quantifying the perceptive difference between two colors is proposed. This method is a complexion of Manhattan's distance and corrects it when the difference between the two colors is not a balanced difference between the three spectral channels.
6. *Guerrero et al. (2012b)*, is a work published on the minutes of the event that took place in Pisa (Italy) during September 2012. This work presents the influence of the pitch angle of the camera on the resolution of the images and therefore on the accuracy of any process done with it.
7. The following is a list of contributions as co-author in proceedings related to the research: *Guerrero et al. (2012c)*, *Montalvo et al. (2012b, 2013b)* and *Guijarro et al. (2013)*.

### **1.5.3. Lectures and Seminars**

- a) Conference titled: "*Stereoscopic vision systems for outdoors robots control and for precision agriculture*" given in the Faculty of Informatic of the Complutense University of Madrid in May 2011. This was held and financed under the partnership agreement of the Fund for International Cooperation in Science and Technology European Union-Mexico (FONCICYT). With the collaboration of the

Polytechnic National Institute and the University of Guadalajara of Mexico, the Free University of Berlin and the Complutense University of Madrid.

- b) Conference titled: “*Practical applications of the artificial vision in outdoor environments*” belonging to the subject “Artificial vision seminars” of the Artificial vision Master of the Rey Juan Carlos I University of Madrid in February 2012.

#### **1.5.4. Radio broadcasts**

Collaboration in the radio program *Sin Distancias* titled “artificial eyes in agriculture” broadcasted the 21<sup>st</sup> of March within UNED programming in RADIO-3 of RNE.

### **1.6. English summary organization**

The English summary of this thesis is organized in six chapters. The present chapter *chapter 1* is an introduction of the works done, the motivation, the objectives, the methodology and the contributions made by the present work. The other five chapters, *chapter 2* to *chapter 6* are dedicated to the algorithms developed along the research period. Each chapter deals with one of the algorithms. Algorithms are widely explained as well as the state of the art. At the end of every chapter results and conclusions are exposed.

# CROP ROW DETECTION IN MAIZE FIELDS INSPIRED ON THE HUMAN VISUAL PERCEPTION

## 2.1. Problem statement

The increasing development of robotics equipped with machine vision sensors applied to Precision Agriculture is demanding solutions for several problems. The robot navigates and acts over a site-specific area of a larger farm (*Davies et al. 1998*), where one important part of the information is supplied by the vision system.

An important issue related with the application of machine vision methods is that concerning the crop row and weed detection, which has attracted numerous studies in this area (*Onyango and Marchant 2003; Tellaeché et al. 2008a,b; Burgos-Artizzu et al. 2009; Sainz-Costa et al. 2011*). This will allow site-specific treatments trying to eliminate weeds and to favor the growth of crops.

The robot navigates on a real terrain presenting irregularities and roughness. This produces vibrations and also swinging in the pitch, yaw and roll angles. Moreover, the spacing of crop rows is also known. Because of the above, the crop rows are not projected on the expected locations in the image. On the other hand, the discrimination of crops and weeds in the image is a very difficult task because their red, green, and blue spectral components display similar values. This means that no distinction is possible between crops and weeds based on the spectral components. Thus, the problem is to locate the crop rows in the image. To achieve this goal, in this work we propose a new strategy that exploits the specific arrangement of crops (maize) in the field and also applies the knowledge of perspective projection based on the camera intrinsic and extrinsic parameters. This method is inspired by the human visual perception and like

humans applies a similar reasoning for locating crop rows in the images, although it exploits the camera system geometry, because it is available. As we will see in the next section, the crop row location is not new and has been considered already in the literature; the method proposed in this work gains advantage over existing approaches because it has been designed to achieve high effectiveness in real-time applications. This makes the main contribution of this work. The method does not include a segmentation step, which is found in most other methods for plant detection. This tries to avoid time consumption as compared to other strategies.

The segmentation step has been replaced by a simple thresholding method, where the threshold is previously established by applying a learning-based fuzzy clustering strategy. Cluster centers for green textures are obtained during an off-line learning phase, and then this knowledge is exploited during the online decision phase. Crop row detection is an easy step where simple straight lines are traced, based on perspective projection, looking for specific pixels alignments defining crop rows.

Moreover it applies down sampling for reducing image sizes. All above steps are oriented to gain time reduction during the computational process. The proposed approach is favorably compared against some existing approaches in both effectiveness and time reducing.

## **2.2. Revision of methods**

Several strategies have been proposed for crop row detection. *Fontaine and Crowe (2006)* tested the abilities of fourth-line detection algorithms to determine the position and the angle of the camera with respect to a set of artificial rows with and without simulated weeds. These were stripe analysis, blob analysis, linear regression, and Hough transform.

### **2.2.1. Methods Based on the Exploration of Horizontal Strips**

*Søgaard and Olsen (2003)* apply RGB color image transformation to grayscale. This is done by first dividing the color image into its red, green, and blue spectral channels and then by applying the well-tested methods to extract living plant tissue described in *Woebbecke et al. (1995)*. After this, the greyscale image is divided into horizontal strips where maximum grey values indicate the presence of a candidate row,

each maximum determines a row segment, and the center of gravity of the segment is marked at this strip position. Crop rows are identified by joining marked points through a similar method to the one utilized in the Hough transform or by applying linear regression. *Sainz-Costa et al. (2011)* have developed a strategy based on analysis of video sequences for identifying crop rows. Crop rows persist along the directions defined by the perspective projection with respect the 3D scene in the field. Exploiting this fact, they apply greyscale transformation, and then the image is binarized by thresholding. Each image is divided into four horizontal strips. Rectangular patches are drawn over the binary image to identify patches of crops and rows. The gravity centers of these patches are used as points defining the crop rows, and a line is adjusted considering these points. The first frame in the sequence is used as a lookup table that guides the full process for determining positions where the next patches in subsequent frames are to be identified. *Hague et al. (2006)* transform the original RGB image to gray scale. The transformed image is then divided into eight horizontal bands. The intensity of the pixels across these bands exhibits a periodic variation, due to the parallel crop rows. Since the camera characteristics, pose and the crop row spacing are known a priori, the row spacing in image pixels can be calculated for each of the horizontal bands using a pinhole model of the camera optics (*Tsai, 1987*). A band-pass filter can then be constructed which will enhance this pattern and has a given frequency domain response. Sometimes horizontal patterns are difficult to extract because crops and weeds form a unique patch.

### **2.2.2. Methods Based on the Hough Transformation**

According to *Slaughter et al. (2011)*, one of the most commonly used machine vision methods for identifying crop rows is based upon the Hough transform (*Hough 1962*). It was intended to deal with discontinues lines, where the crop stand is incomplete with gaps in crop rows due to poor germination or other factors that result in missing crop plants in the row. It has been intended for real-time automatic guidance of agricultural vehicles (*Marchant 1996; Hague et al. 1997; Astrand and Baerveldt 2005; Leemans and Destain 2006*). It is applied to binary images, which are obtained by applying similar techniques to the ones explained above, that is, RGB image transformation to grayscale and binarization (*Tellaeché et al. 2008a,b; 2011*). *Gee et al. (2008)* apply a double Hough transform under the assumption that crop rows are the

only lines of the image converging to the vanishing point, the remainder lines are rejected, additional constraints such as inter-row spacing and perspective geometry concepts help to identify the crop rows. It is required to determine the threshold required by the Hough transform to determine maximum peaks values (*Jones et al. 2009a,b*) or predominant peaks (*Rovira-Mas et al. 2005*). Depending on the crop densities, several lines could be feasible, and a posterior merging process is applied to lines with similar parameters (*Tellaeche et al. 2008a,b, 2011*). *Ji and Qi (2011)* report that Hough transform is slow due to the huge computation; they propose a randomized Hough transform to reduce computational time. Some modifications have been proposed to improve the Hough transformation such as the one proposed in *Asif et al. (2010)*, which apply the Hough only to those points which are edge points along the crops. But this requires the application of techniques for edge extraction. Also the randomized Hough transformation has been proposed with this goal (*Ji and Qi, 2011*). It is intended to avoid redundant computations in the Hough transform. It operates iteratively by randomly sampling a set of points to compute a single localization in the Hough space. Since two pixels are trivially collinear, the parameters of the line on which they lie can be estimated. These parameters are used to increment the accumulator cell in the Hough space. In summary, the Hough transform is computationally expensive and the randomized Hough transform requires selecting pairs of points to be considered as a unique line, that is, pairs of points belonging to a crop row. If we apply this technique in images where edge points have been extracted, the selection of those pairs becomes more complex. Furthermore, the computational cost of Hough-based algorithms is very sensitive to the image resolution after down-sampling, but also when weeds are present and irregularly distributed, this is our case, this is could cause the failure detection. Moreover, as the weed density increases the crop row detection becomes more and more difficult.

### **2.2.3. Vanishing Point Based**

*Pla et al. (1997)* propose an approach which identifies regions (crops/weeds and soil) by applying color image segmentation. They use the skeleton of each defined region as a feature to work out the lines which define the crop. The resulting skeletons of each region can be used as curves which define the underlying structure of the crop and to extract the straight lines where the plants and soil rows lie. Segments in the

skeletons are defined as chains of connected contour points and they must be of a defined length. This allows selecting candidate lines for crop row detection, among all candidates the ones that meet the vanishing point. The vanishing point is detected using previous information about the vanishing point found in the previous images, performing a sort tracking on the vanishing point. This process is highly dependent on the skeletons, which are not always easy to extract and isolate, particularly considering that crops and weeds patches appear overlapped among them.

#### **2.2.4. Methods Based on Linear Regression**

Some of the techniques above apply this approach. *Billingsley and Schoenfisch (1995,1997)* reported a crop detection system that is relatively insensitive to additional visual “noise” from weeds. They used linear regression in each of three crop row segments and a cost function analogous to the moment of the best-fit line to detect lines fitted to outliers (i.e. noise and weeds) as a means of identifying row guidance information. As mentioned above, *Søgaard and Olsen (2003)* also apply linear regression, which is a feasible approach if weed density is low and pixels belonging to crop rows are well separated. Otherwise it is highly affected by pixels belonging to weeds because of their strong contribution to line estimation.

#### **2.2.5. Stereo Based Approaches.**

*Kise et al. (2005)* and *Kise and Zhang (2008)* developed a stereovision-based agricultural machinery crop row tracking navigation system. Stereovision processing is used to determine 3D locations of the scene points of the objects of interest from the obtained pair of images. Those 3D positions, determined by means of stereovision disparity computation, provide the base information to create an elevation map which uses a 2D array with varying intensity to indicate the height of the crop. This approach requires crops with significant heights with respect the ground. Because in maize fields, during the treatment stage, the heights are not relevant, it becomes ineffective in our application. *Rovira-Mas et al. (2008, 2010)* have applied and extended stereovision techniques to other areas inside Precision Agriculture. Stereo-based methods are only feasible if crops or weeds in the 3D scene display a relevant height and the heights differ in both kinds of plants.

### **2.2.6. Methods Based on Blob Analysis**

This method finds and characterizes regions of contiguous pixels of the same value in a binarized image (*Fontaine and Crowe 2006*). The algorithm searches for white blobs (inter-row spaces) of more than 200 pixels, as smaller blobs could represent noise in the crop rows. Once the blobs are identified, the algorithm determines the angle of their principal axes and the location of their center of gravity. For a perfectly straight white stripe, the center of gravity of the blob was over the centerline of the white stripe, and the angle was representative of the angle of the inter-row spaces. The algorithm returned the angle and center of gravity of the blob closest to the center of the image. Identification of blobs in images infested with weeds in maize fields becomes a very difficult task, because weeds and crops under overlapping in localized areas produce wide blobs.

### **2.2.7. Methods Based on the Accumulation of Green Plants**

*Olsen (1995)* proposed a method based on the consideration that along the crop row an important accumulation of green parts in the image appears. The image is transformed to gray scale, where green parts appear clearer than the rest. A sum curve of gray levels is obtained for a given rectangular region exploring all columns in the rectangle. It is assumed that vertical lines follow this direction in the image. Images are free of perspective projection because they are acquired with the camera in orthogonal position. A sinusoidal curve is fitted by means of least squares to the sum curve previously obtained. Local maxima of the sinusoid provide row centers locations. This is a simple and suitable method, which can be still simplified but it is not of our interest because of the fact that the images we work with are taken from the tractor under perspective projection but not orthogonal projection. In this work we exploit the idea of green plant accumulation under a simpler strategy.

### **2.2.8. Methods Based on Frequency Analysis.**

Because crop rows are vertical in the 3D scene, they are mapped under perspective projection onto the image displaying some behavior in frequency domain. *Vioix et al. (2004)* exploit this feature and apply a bidimensional Gabor filter, defined as a modulation of a Gaussian function by a cosine signal. The frequency parameter

required by the Gabor filter is empirically deduced from the 2D-Fast Fourier *Transform* *Bossu et al. (2006)*.

*Bossu et al. (2009)* apply wavelets to discriminate crop rows based on the frequency analysis. They exploit the fact that crop rows are well localized in the frequency domain; thus selecting a mother wavelet function with this frequency the crop rows can be extracted. In maize fields where the experiments are carried out, crops do not display a clear frequency content in the Fourier space, therefore the application of filters based on the frequency becomes a difficult task.

### **2.3. Motivational research and design of the proposed strategy**

Our work is focused on crop row detection in maize for specific treatments requiring discrimination among crops and weeds. This means that crop rows must be identified and located in the image conveniently. Some of the requirements proposed by *Astrand (2005)* and reported in *Slaughter (2008)* for guidance systems can be considered for crop row detection because the problem is essentially similar. Therefore, our system is designed to be:

- a) Able to locate crop rows with the maximum accuracy as possible.
- b) Able to work on real-time.
- c) Able to work on sown crops, not manually planted, which means that weeds and crops grow simultaneously displaying, at the early growth stage of the treatment, similar heights and also similar spectral signatures. This means that discrimination between crops and weed cannot be made by height or spectral signatures only.
- d) Able to work when plants are missing in the row.
- e) Able to work when there is high weed pressure.
- f) Able to work under different weather (luminance) conditions.
- g) Able to locate crop rows with the least assumptions and constraints.

The aim of this study is to present a general method for identifying crop rows in maize fields from the images. We exploit the advantages of the existing methods introduced above, extracting the main ideas, and design a new strategy for crop row detection inspired on the human visual perception abilities which is able to work in real time.

This method is also dedicated to be applied in maize with crop row spacing and also to deal with and without seedling spacing. It is summarized in the two main steps as follows.

- a) Segmentation of green plants (crop and weeds).
- b) Crop row identification.

## **2.4. Materials and Methods**

### **2.4.1. Images**

The images used for this study correspond to maize crops. They were captured with a Canon EOS 400D camera during April/May 2011 in a 1.7 ha experimental field of maize in La Poveda Research Station, Arganda del Rey, Madrid at the Spanish Research Council facilities. All acquisitions were spaced by five/six days, that is, they were obtained under different conditions of illumination and different growth stages. The digital images were captured under perspective projection and stored as 24bit color images with resolutions of 5 MP saved in RGB (red, green, and blue) color space in the JPG format. The images were processed with Matlab R2009b (*The Mathworks 2014*) under Windows 7 and Intel Core 2 Duo CPU, 2.4 GHz, 2.87 GB RAM. A set of 350 images were available for processing.

With the aim of testing the robustness and performance of the proposed approach, we have worked with images captured under different conditions, including different number of crop rows; these conditions have been identified in the real fields as possible and also those that could cause problems during the detection process in normal operation. The following is a list of representative images from the set of available images, illustrating some of such conditions:

- a) Different brightness due to different weather conditions, *figures 2.1(a)* and *2.1(b)*;
- b) Different growth stages, *figures 2.2(a)* and *2.2(b)*;
- c) Different camera orientations, that is, different yaw, pitch and roll angles, and heights from the ground, *figures 2.3(a)* and *2.3(b)*;
- d) Different weeds densities, *figures 2.4(a)* and *2.4(b)*.

### **2.4.2. Image Segmentation: Green Plants Identification**

For real-time applications is of great relevance to simplify this process as much as possible. Instead of using vegetation indices (*Woebbecke et al. 1995* and *Ribeiro et al. 2005*), which require an image transformation from RGB color space to gray scale, we used a learning-based approach with the goal of obtaining the percentage of the green spectral component with respect to the remainder, which allowed us to consider a pixel belonging to a green plant.

This relative percentage is intended to deal with illumination variability so that it determines relative values among the three spectral RGB components that identify green plants. This is carried out by applying a fuzzy clustering approach. Under this approach there is a learning phase which is applied to during offline activity for computing the relative percentage or threshold and a decision phase where the threshold is applied without additional computation.

The learning phase was designed as follows. From the set of available images we randomly extracted  $n$  training samples, stored in  $X$ , that is,  $X = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathfrak{R}^d$ , where  $d$  is the data dimensionality. Each sample vector  $\mathbf{x}_i$  represents an image pixel, where its components are the three RGB spectral components of that pixel at the original image location  $(x,y)$ . This means that in our experiments the data dimensionality is  $d=3$ .

Each sample is to be assigned to a given cluster  $w_j$ , where the number of possible clusters is  $c$ , that is,  $j=1, 2, \dots, c$ . In the proposed approach  $c$  is set to 2 because we were only interested in two types of textures, that is, green plants (crop/weeds) and the remainder (soil, debris, stones).

The samples in  $X$  are to be classified based on the well-known fuzzy clustering approach that receives the input training samples  $x_i$ , and establishes a partition, assuming the number of clusters  $c$ , is known. The process computes for each  $x_i$  at the iteration  $t$  its degree of membership in the cluster  $w_j(\mu_{ij})$  and updates the cluster centers  $v_j$  as follows (Zimmermann 1991):

$$\mu_i^j(t+1) = \frac{1}{\sum_{r=1}^c \left( \frac{d_{ij}(t)}{d_{ir}(t)} \right)^{\frac{2}{b-1}}} \quad (2.1)$$

$$v_j(t+1) = \frac{\sum_{i=1}^n [\mu_i^j(t)]^b x_j}{\sum_{i=1}^n [\mu_i^j(t)]^b} \quad (2.2)$$

$d_{ij}^2 \equiv d^2(\mathbf{x}_i, \mathbf{v}_j)$  is the squared Euclidean distance. The number  $b$  is called the exponential weight (Bezdek 1981 and Duda et al. 2000),  $b > 1$ . The stopping criterion of the iteration process is achieved when  $\|\mu_i^j(t+1) - \mu_i^j(t)\| < \varepsilon \quad \forall ij$  or a number  $t_{max}$  of iterations is reached.

The method requires the initialization of the cluster centers, so that equation (2.1) and (2.2) can be applied at the iteration  $t = 1$ . For this purpose, we applied the pseudorandom procedure described in Balasko et al. 2008.

- a) Perform a linear transform  $Y = f(X)$  of the training sample values so that they range in the interval  $[0, 1]$ .
- b) Initialize  $v = 2DM \circ \mathbf{R} + D\mathbf{m}$ , where  $\mathbf{m}$  is the mean vector for the transformed training samples values in  $Y$  and  $\mathbf{M} = \max(\text{abs}(Y - \mathbf{m}))$ , both of size  $1 \times d$ ;  $D = [1 \dots 1]/T$  with size  $c \times 1$ ;  $\mathbf{R}$  is a  $c \times d$  matrix of random numbers in  $[0, 1]$ ; the operation  $\circ$  denotes the element by element multiplication.



*Figure 2.1: Different brightness due to different weather conditions: (a) darker; (b) clearer.*



*Figure 2.2: Different crop growth stages: (a) low; (b) high.*



*Figure 2.3: Different yaw, pitch and roll angles, and heights from the ground.*



*Figure 2.4: Different weed densities: (a) low; (b) high.*

When the learning process is finished we obtain two cluster centers  $\nu_1$  and  $\nu_2$  associated to clusters  $w_1$  and  $w_2$ . Without loss of generality, let  $\nu_1 \equiv \{\nu_{1R}, \nu_{1G}, \nu_{1B}\}$  the one associated to the green plants. It is a 3-dimensional vector where its components  $\nu_{1R}$ ,  $\nu_{1G}$ , and  $\nu_{1B}$  represent the averaged values for the corresponding RGB spectral components; thus the threshold value for discriminating among green plants and the remainder ones is finally set to  $T_G = \nu_{1G}/(\nu_{1R} + \nu_{1G} + \nu_{1B})$ . Once  $T_G$  is available, the green parts on the images are identified assuming the corresponding RGB pixels contain the  $G$  spectral value greater than  $T_G$ . Therefore, during the online identification process only is required the logical comparison.

### 2.4.3. Crop Row Identification

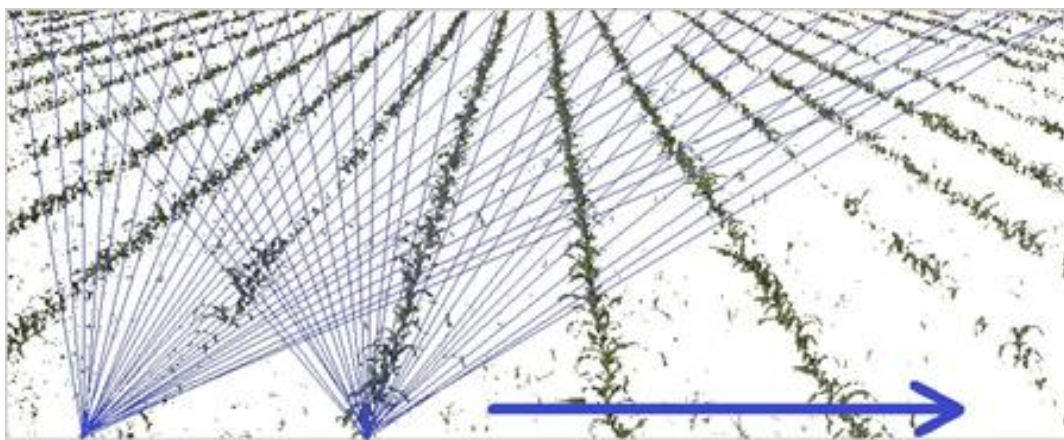
Once green parts were extracted in the image, next step was crop row identification. For such purpose we make use of the following constraints, based on the image perspective projection and the general knowledge about the maize field.

1. The number of crop lines  $L$ , to be detected is known and also the approximate  $x$  position or image column at the bottom of the image where every crop line starts. This assumption is based on the system geometry and the image perspective projection.
2. We are going to detect crop lines that start from the bottom of the image and end at the top of the image. Lines starting from both left and right sides of the image and vanishing at the top are rejected. This is because image geometry allows considering this situation. An extension of this algorithm could be done to detect those lines with its corresponding computing time cost.
3. All the images have been acquired with a camera onboard a tractor and pointing in the same direction as the crop lines, therefore images are mapped under in perspective projection and the crop lines converge in the vanishing point. This constraint is inspired by methods based on the vanishing point, as described in the introduction. Though crop lines are parallel, distance between crop lines seems to be greater at the bottom of the image than at the top, due to perspective. This algorithm works considering that crop lines are going to have that appearance in the image

with a range of tolerance that can be adjusted depending on the stability of the tractor and the evenness of the ground. We assume that crop lines starting from the left bottom of the image have a clockwise slope and lines starting from the right bottom of the image have an anticlockwise slope. The bigger the range of tolerance the higher the computing time. In this work we have used a 15% of tolerance which means that crop lines may vary from one image to the next one a 15% of the width of the image.

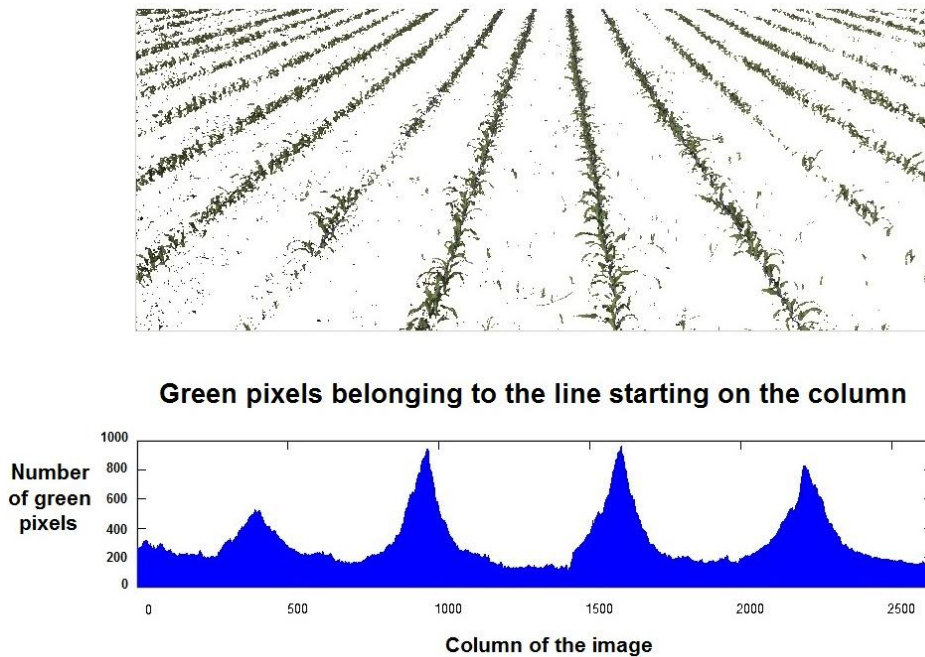
The algorithm works as follows:

1. From every pixel in the bottom row we trace all the possible lines starting on that pixel and ending on every pixel of the top line, that is, if the image has  $N$ -columns, we will trace  $N^2$  lines from every pixel of the bottom row, which means that we finally trace  $N^2$  lines. This number of calculated lines is the highest number of lines in case we make no constraints. Nevertheless, as we will see in *step 2* and *step 3* important constrains can be applied to reduce this number. *Figure 2.5.* shows the beam of lines traced for two pixels placed at the bottom row of the image. For illustrating purposes we have traced broad beams of lines, but the number of lines to cope with all possible situations, but this number could be considerably reduced by applying previous knowledge, like the slope. This is applied in this work as described below, reducing the computational cost.



**Figure 2.5:** Lines traced from every pixel of the bottom row to every pixel of the top row in the image.

- For every traced line starting on a pixel of the bottom row, we count the number of green pixels that belong to that line. This is possible because the image has been already segmented and pixels belonging to green parts have been identified. Now, the line with the highest number of green pixels is the candidate line to represent the crop row for that pixel of the bottom row.



**Figure 2.6:** Number of green pixels found for the best line of every pixel of the bottom line.

- We repeat the same procedure for all the pixels of the bottom row, and finally we obtain  $c$  candidates lines, that is, one for every pixel of the bottom row. As we can see in *figure 2.6*, every pixel location in the bottom row of the image has a value for its best line. Those values become higher as the represented line approaches the real crop line. They are the peaks in the lower part in *figure 2.6*.
- Since we know the number of crop lines to be detected and also the pixel where they roughly start at the bottom row, we can choose the closest and highest values to those pixels, which are identified by peaks in the figure. In the figure shown we assume the number of crop rows to be detected is four; so we look for four peaks that are conveniently spaced because of the crop rows arrangement in maize fields and also based on camera system geometry. This idea is inspired in methods based on the accumulation of green plants, described in the introduction.

The following are two considerations that can be applied to speed up the computational process from the point of view of a real-time application.

1. Because of the perspective of the image it is not necessary to trace all the lines to the top row (as mentioned in *step 1*) but only those whose slope is according to what we expect. That is, if we are dealing with left pixels of the image we would trace only lines with a slope clockwise and without reaching the end of the right side of the image. For right pixels we would search for anticlockwise slopes starting from the right side and without reaching the left side of the image. This idea is based on the vanishing point concept and applies in some approaches as described in the introduction.
2. In addition to it, it is not necessary to trace lines pixel by pixel. Depending on the image resolution a “pixel step” can be used without affecting final result and reducing considerably the computational cost.

Notice that there are some values to be adjusted before the algorithm runs. These values depend on the images we are dealing with and on the stability of the camera. The higher the image resolution, the higher the “pixel step” for lines calculation. Furthermore, the better the stability of the tractor the thinner the range of pixels of the top row for tracing lines.

## 2.5. Results.

Our proposed Crop Row Detection (CRD) method consists on a first stage or learning phase where the threshold  $T_G$  is obtained for segmenting green plants. With such purpose we have processed 200 images, selected from the set of 350 images available, from which 40.000 training samples have been randomly extracted. The selected images cover the broad range of situations, that is, different number of rows, weather conditions, weed concentrations and growth states according to *figures 2.1 to 2.4*.

With these training samples, we apply the fuzzy clustering procedure described in *section 2.4.2 “Image Segmentation: Green Plants Identification”*, from which we compute two cluster centers identifying both green plants ( $\mathbf{v}_1$ ) and soil or other components ( $\mathbf{v}_2$ ). Our interest is only focused on segmenting green plants, therefore,

from  $\mathbf{v}_1$  we compute the threshold  $T_G$  defined in *section 2.4.2 “Image Segmentation: Green Plants Identification”* as the percentage of the green component in  $\mathbf{v}_1$ , that is,  $T_G = 0.37$ . This is the threshold finally used for image segmentation.

*Table 2.1* displays both cluster centers  $\mathbf{v}_1$  and  $\mathbf{v}_2$  and the percentage of the highest value in the spectral components associated to each cluster center. As we can see green and soil pixels can be identified by the corresponding percentages, each one applied over the green and red spectral components. This was the general behavior observed for the set of images analyzed.

	Spectral component values	Percentage of the highest spectral component
$\mathbf{v}_1$ (green plants)	{13780 140.68 106.07}	0.37 (green)
$\mathbf{v}_2$ (soil and other components)	{188.49 177.71 153.53}	0.36 (red)

**Table 2.1:** Percentage of the green spectral component for green plants and for other components (soil, debris, stones).

As mentioned during the introduction, the Hough transform has been applied in several methods for crop row detection, hence we compare the performance of our CRD approach against the Hough (HOU) transform. We have applied identical conditions to the Hough transform than the ones applied in our CRD approach, so it works in terms of comparability; they are synthesized as follows.

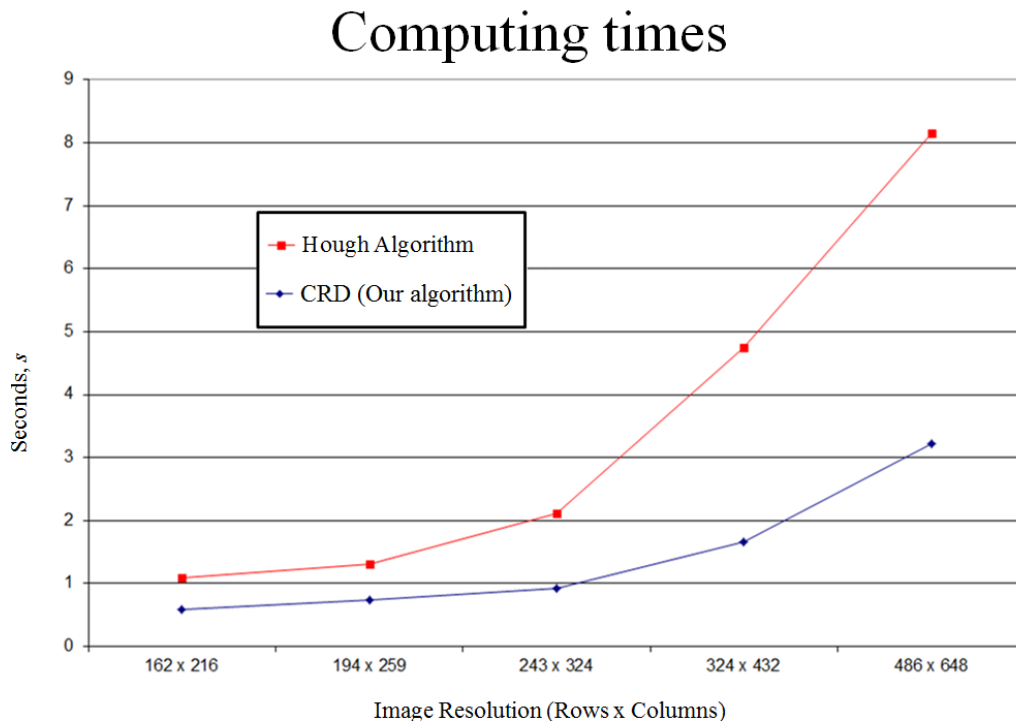
1. Search for lines arising from the bottom of the image and ending at the top, that is, suspicious useless lines are not explored.
2. Only are allowed lines with slopes close to the ones expected at each side of the image. Horizontal lines and many others that do not meet the above are rejected.
3. The Hough transform is implemented to work under the normal representation, polar coordinates (*Gonzalez et al. 2009*), with unit increments in the parameter representing the angle.

The comparison is established in terms of effectiveness and processing times. The effectiveness is measured based on the expert human criterion, where a line, which has been detected, is considered as correct if it overlaps with the real crop row alignment. Over the set of 350 images analyzed, we compute the average percentage of

coincidences for both CRD and HOU. Also, because the main goal of the proposed approach is its profit for real time applications, we measure computational times. Also, with the goal of real-time, we have tested these performances for different image resolutions. As we can see image resolutions differ from the ones in the original images, these resolutions have been obtained by applying a down-sampling process to the original image. This is intended under the idea that it is possible to reduce the image dimension retaining the main information without affecting the effectiveness and reducing the processing time. *Table 2.2* displays the results. The first column contain different image resolutions, which are obtained by selecting large regions of interest in each image, with horizontal and vertical sizes of  $1940 \times 2590$  pixels, and these regions contains different number of crop rows under different configurations provided by the images displayed in *figures 2.1.* to *2.4.* These large regions are split by 10, 8, 6, and 4, which are, respectively the ones represented in *table 2.2.* We have chosen this set of values because with them we obtain similar performances in terms of effectiveness with acceptable processing times. The effectiveness for higher resolutions is similar, but the processing times increase considerably. Below the lower resolution, the effectiveness decreases considerably.

Image resolution (pixels)	Percentage of effectiveness		Processing time(seconds)	
	HOU	CRD	HOU	CRD
<b>162 x 216</b>	86.3	97.1	1.088	0.580
<b>194 x 259</b>	89.4	97.3	1.305	0.737
<b>243 x 324</b>	89.1	97.3	2.120	0.928
<b>324 x 432</b>	90.9	97.4	4.752	1.667
<b>486 x 648</b>	91.1	97.5	8.153	3.216

**Table 2.2:** Performances of HOU and Crop Row Detection, CRD approaches measured in terms of percentage of effectiveness and processing times.



**Figure 1.8:** Computing times (seconds) against the different image resolutions.

The second and third columns contain the percentage of effectiveness and columns fourth and fifth the processing times measured in seconds. All these measurements represent average values over the set of 350 images processed. For clarity, *figure 2.7* represents the processing times in *table 2.2.*, for the four resolutions studied. From results in *table 2.2* we can infer that CRD outperforms HOU in terms of effectiveness, with a near constant value regardless the image resolution. With lower resolutions, that is, with image divisions above 12, this percentage decreases drastically, achieving values below 85%. This is because for low resolutions some important information in the images is lost. Thus, from values in *table 2.2* and because the processing time is lower with small image resolutions, from a real-time point of view, a suitable resolution with acceptable performance is the lowest, that is, the one for  $162 \times 216$ . From *figure 2.7*, one can see that the increasing of time is not linear. For resolutions above  $243 \times 324$  time differences are more pronounced.

The worst performance obtained for HOU can be explained because crops and weeds concentration produces a high density of values, representing peaks, in the cell accumulator. These values do not display a high clear value, theoretically representing a unique crop line. Moreover, the absolute maximum value around the expected crop line

most times does not represent the correct line. Thus, it is necessary to define a patch selecting different high peak values for each expected line, which are averaged, to obtain the final value. Because this patch has not clear limits, its selection becomes a difficult task and errors in the selection produce errors in the crop row localization, which explain the worst performance of HOU against CRD.

## **2.6. Conclusions.**

We developed a new method for crop row detection that improves Hough-based methods in terms of effectiveness, robustness and computing time. The goal is its application to real-time implementations.

Furthermore, our approach has been proved to be robust enough to different images typologies.

The proposed method is robust enough to work in images under perspective projection. It can detect any number of crop lines with any slope converging in a vanishing point. It works with either high or low image resolutions, and no matter the presence of high weeds concentration.

Future works must be oriented toward weeds detection by establishing cells around crop lines and calculating the percentage of greenness of every cell. This should be intended for posterior actuations to kill weeds.



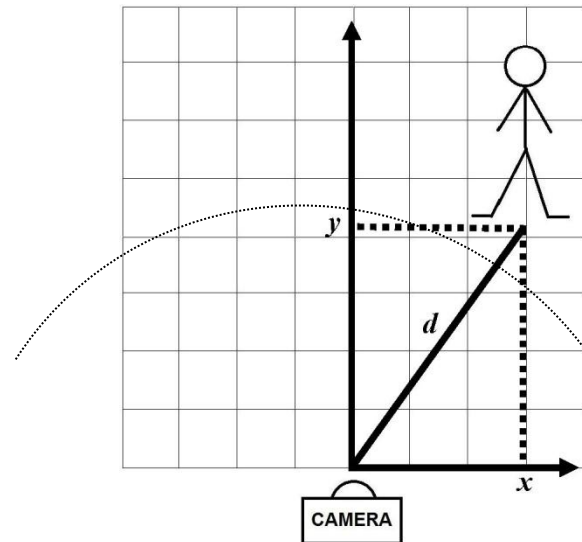
## AUTONOMOUS ROBOTS POSITIONING IN WATER SURFACE FROM A SINGLE IMAGE

### 3.1. Introduction

There are many depth computation techniques, based on computer vision, to determine the distance of an object in the scene 3D to a reference system. Most of them are focused on the computation of distances without considering the object location. This means that different locations provide the same distance. *Figure 3.1* displays a pedagogical example where the object located at position  $(x,y)$  is also placed at the distance  $d$ , but this distance is the same for all points in the circumference with radius  $d$  and center in the camera.

Autonomous navigation of unmanned boats requires the knowledge of their locations with respect a reference system for control purpose. The proposed method is specifically designed to provide the required boat 3D locations based only on the computation of their corresponding row and column pixels coordinates on a single image.

Additionally, two important advantages of the proposed approach are that it does not requires knowledge about camera's intrinsic and extrinsic parameters and also that avoids the use of more complex, sophisticated and expensive systems onboard the boats, such as GPS with communication systems, requiring power supplies or additional sources of energy. Motivational research is supported by the above advantages, where the main contribution of this method is the computation of both  $x$ - $y$  coordinates.



*Figure 3.1:  $x$ - $y$  coordinates and distance of an object from a reference system whose origin is set on the camera.*

### 3.2 State of the art

In recent years, several methods for depth estimation have been proposed. They are classified in a field called: 3D Computer Vision. The main goal of this area is the design of strategies to determine the position of objects in the scene 3D. These techniques may be grouped into two categories: active and passive methods (*Kumar-Mada et al. 2002*). Active techniques apply energy through a structural source (laser, radar, patterns of light) and receive responses through a detector. Based on the nature of the interaction with the surfaces of the objects, the reflected energy is used for recovering the 3D structure in the scene. It is well known that in outdoor environments there are serious limitations and conflicts with variations in object reflectance, color and range. Alternatively, passive vision techniques do not apply specific structured source of energy to build images avoiding interferences with the environment. This work is concerned with passive methods. Most works have addressed the topic of 3D determination with passive methods; they are grouped under the generic expression of shape from  $X$ , where  $X$  can be (*Trucco and Verri, 1998*): a) stereopsis (*Scharstein and Szeliski, 2002*); b) motion, where the object is moving and the camera is static (*Buch et al., 2009 and Toshev et al., 2009*) or the camera moving into the 3D space (*Leibe et al., 2007*); c) shading (*Horn 1986; Zhang et al., 1999; Prados and Faugeras, 2006*); d) texture based on repetition of objects or patterns in the scene (*Payet and Todorovic, 2011; Lobay and Forsyth, 2006*); e) focus/defocus: based on a well-focused object which is mapped exactly on the image plane (*Krotkov, 1987; Subbarao et al., 1993*;

*Subbarao and Surya, 1994*); f) depth estimation with a single camera, *Saxena (2008)* applied supervised learning; *Kaushik (2012)* determines relationships between distances of objects in the scene and their corresponding pixels in the image.

### **3.2.1. Stereopsis**

Based on two cameras separated a base-line, a 3D point in the scene is mapped into different row  $r$ , and column  $c$ , pixel coordinates on each image. A matching procedure is applied to identify the corresponding point in both images. Once the correspondence is established, the 3D pose is obtained by triangulation. Matching is a critical issue in stereovision, which becomes ineffective when large homogeneous areas appear in the scene. Under these situations the matching becomes unfeasible. This is the case in the proposed application where the water surface displays high texture homogeneity, providing false matches that can be confused with objects in the 3D scene. From a certain distance onwards boats are mapped as small patches without a clear shape, making difficult its identification based on pattern recognition techniques. Moreover, the continuous movement of boats over the water surface requires that both cameras in the stereovision system are completely synchronized for capturing the images simultaneously and also that both cameras provide similar intensity levels. Depth estimation is commonly performed by stereopsis based techniques (*Scharstein and Szeliski 2002*). In recent years several modifications of the classical methods have been developed and applied in metrology to determine features of the objects in images, e.g. height, volume or distance. In *Chen et al. (2006)*, is proposed a robust method to measure object height based on two cameras, it computes the height using different planes of the image and geometric transformations. On the other hand, there are algorithms that only obtain relative depths and disparity maps without object localization in the scene (*Aziz and Mertsching, 2010; Daolei and Lim, 2011; Agrawal et al., 2007*). Other implementations combine algorithms in order to reduce the computational time required to obtain depth maps (*Shan-Shan et al., 2011; Birchfield S. and Tomasi, 1998*). Most of such approaches perform depth and object features estimation under acceptable computation conditions. Nevertheless, they still need intrinsic and extrinsic parameters for both cameras, they don't provide numeric values of the distance between cameras and objects and only have been tested in indoor environments with controlled illumination images. In outdoor environments, agriculture, stereovision has been applied successfully (*Rovira-Mas et al., 2008*).

### 3.2.2. Motion

Moving objects produce changes in subsequent images displaying areas which change with respect to the background. These methods compute relative motion between camera and 3D scene with small changes. These changes are detected by differential analysis in both spatial and temporal dimensions (optical flow). 3D reconstruction becomes a complex task because the differentiation requires small movements with few and short features in objects (*Ozden et al. 2010*). In our scenario, boats are in continuous movement producing continuously arbitrary displacements which are not necessarily small. One way to acquire the 3D data is to determine the direction of translation through approximate motion parallax. Depths can be obtained by least squares estimations from the optical flow (*Cantzler 2012*). Because of the above mentioned continuous displacements, recovery of motions and 3D structure is highly affected by this effect, becoming ineffective in our outdoor environments for boats location.

### 3.2.3. Shading

These methods apply patterns of lights and the corresponding shades to infer the shape. It is required the previous knowledge of angles of incidence of the light and also a surface shape estimation. Shading analysis fails when changes in the reflectance occur; this is quite common in outdoor environments. In our scenario this becomes practically unfeasible because the illumination appears scattered and the surface is plane. Objects on the water surface, at certain distances, do not display sufficient shades for 3D recovering and do not either show known surface shapes. Moreover, only relative depths are obtained without relation to the world reference system.

### 3.2.4. Texture

Methods based on texture analysis exploit the existence of regular repetitions of an element or pattern, called surface texel, in the surface of the structures in the image. Thus, based on the distortion of these texels, once they are mapped over the image, the 3D structure can be inferred by analyzing the slant of such texels when they are mutually compared or by analyzing the distortion of patterns due to perspective projection, e.g.: circumferences are mapped as ellipses. This last assumption requires that texture features are constant within regions of uniform texture. In our water

scenario, repetitions of objects or constant features do not appear and the water surface is large with only small boats present displaying several faces, thus texture based approaches become useless. Additionally, it is required knowledge of the camera's intrinsic parameters.

### **3.2.5. Focus/Defocus**

It is based on the assumption that well-focused objects in the image are mapped on the image plane, thus if we know the focal length, a simple triangulation allows determining the 3D position. This approach requires adjusting the focal length to determine well focused areas, but in homogeneous surfaces this becomes impractical. Also this requires a motorized system for lens adjusting requiring special calibration to achieve sufficient accuracies for recovering the 3D structure (*Tao et al. 2010*). Different techniques are available for computing measures of focusing, most of them based on edge sharpness. In our scenario the large water surfaces do not display edges and there are not parts to be checked for focusing. Moreover, boats objects, placed at relatively high distances, do not display edges which can be detected with the required precision for focusing purposes.

### **3.2.6. Depth estimation with a single camera.**

Several approaches under different point of view can be considered here. (*Saxena et al. 2008*) apply supervised learning-based methods, where relationships are established between structures in the image and depths in the scene based on the availability of ground-truths. These relations are later applied for new incoming images from similar scenarios. In our scenario, water planar surface appears as a large structure at very different levels of depth, this makes unfeasible the application of this approach.

*Wang et al. (2005)* and *Criminisi et al. (2000)* consider vanishing points in the image and a reference plane in the scene to compute the projection matrix, from which object's heights are derived. In the same way (*Fernandes and Neves 2006*) proposed to measure distances based on a single camera with a specific orientation that is capable to estimate the distance between the object and the camera using specific calibration and trigonometric identities. These approaches are based on the previous knowledge of heights in real objects. *Hsu et al. (2009)* compute depths based on camera displacements, where a relation is established between camera displacement and the

variation of pixels in the image for the object under consideration. In the above approaches objects remain still during depth computation and located to relatively small distances. This differs from our application where boats are moving and also are located at distances relatively high.

*Kaushik (2012)* obtains samples of reference objects from different images. Positions of the reference objects are known in the 3D scene and also pixel *y-coordinate* in the captured images. An  $n$ -degree polynomial is estimated, which describes the *y-coordinate* in the 3D scene, as a function of pixels where objects are located in the image.

*Salvi et al. (2002)* perform a revision of camera calibration methods, including linear techniques which compute the direct transformation matrix through least squares between the images coordinates, in pixels and the real ones, in length units. (*Hall y col., 1982; Faugeras y Toscani, 1986; Ito, 1991*). Because the lens introduces non-linear radial distortions, these methods are not as appropriate as expected.

The aim of this study is to present a general method valid for positioning boats in water planar surfaces at relatively large distances (up to around 150 meters). From the point of view of texture image analysis, water surfaces display a high degree of homogeneity. Because of the nature of our problem and based on limitations of above strategies, we propose a method based on a single image which is independent of extrinsic and intrinsic camera parameters, so that calibration problems in this complex scenario can be avoided. Some of the strategies described above provide only relative depths of objects in the scene in relation to other objects, i.e. they are inappropriate for our problem. Stereo-based and some single image-based methods provide only distances but not  $x$ - $y$  3D locations required for boat positioning. The closest approach to the one proposed in this work is the method proposed on *Kaushik (2012)* but insufficient for positioning as it only provides the *y-coordinate*. We also estimate mathematical functions to establish relations between real coordinates and pixels of objects in the image. These functions, not polynomials, allow estimating, with sufficient accuracy for control purposes, both *x-coordinate* and *y-coordinate* involving pixel rows and columns of objects in the image. The main contribution of our algorithm is the computation of both  $x$  and  $y$  real coordinates, from which we can compute distances like some of the above methods, including stereopsis, at the same time we are able to

position boats in the scene, which was the key objective. Specifically, computation of the  $x$  coordinate from  $y$  is the principal finding. Additionally, as mentioned before, with the proposed approach we avoid calibration of the single camera and do not require knowledge about camera system geometry nor intrinsic parameters. This passive method achieves equivalent positioning results to the ones provided by GPS active system, which is used for comparative purposes.

### 3.3. Method description

As mentioned before, we propose a method to estimate the real 3D  $x$ - $y$ - $z$  coordinates in a scene corresponding to a pixel in the image by knowing only the row and the column where the pixel is placed in the image. The coordinates are estimated with respect to a world reference system  $OXYZ$ , that in our case, we set on the camera with the origin  $O$  coincident with the center of the image plane and the  $Y$ -axis coincident with the optical axis in the camera and perpendicular to the image plane, the  $Z$ -axis is perpendicular to the ground and it is not needed for this method, *figure 3.2(a)*. We have tested this method in a reservoir with images of a boat taken from a camera placed on the shore at a certain height from water surface. Initially, to obtain the parameters involved in the mathematical functions to be estimated, a manned boat was equipped with a differential GPS so that its real position ( $GPS$  coordinates) was known at every time. The images captured for this estimation were taken while the boat was moving on different positions inside the field of view of the camera and its  $GPS$  coordinates were logged.

Since GPS coordinates are not the same as coordinates from our system of reference, we have to make a change of coordinates from GPS axes to ours, which is placed on the camera. From now on when we refer and show GPS coordinates we assume this change of coordinates done.

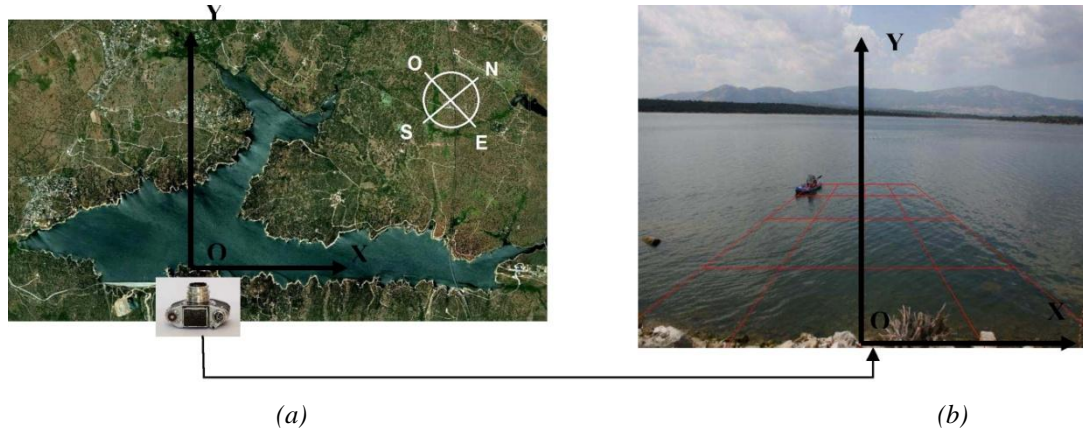
Our method consists of two stages:

a) Estimation of two mathematical functions linking pixels rows  $r$ , and columns  $c$ , in the image and real 3D  $x$ - $y$  coordinates in the scene.

b) Once the above two functions are estimated the boat navigates freely again and its positioning on every new image taken is provided by the two estimated functions and the results are compared with the GPS coordinates also logged.

The following two requirements are needed to apply the proposed method:

- a) The camera must be fixed permanently, during both stages.
- b) Boats must be placed on planar surfaces, i.e.: Water surface in our approach.



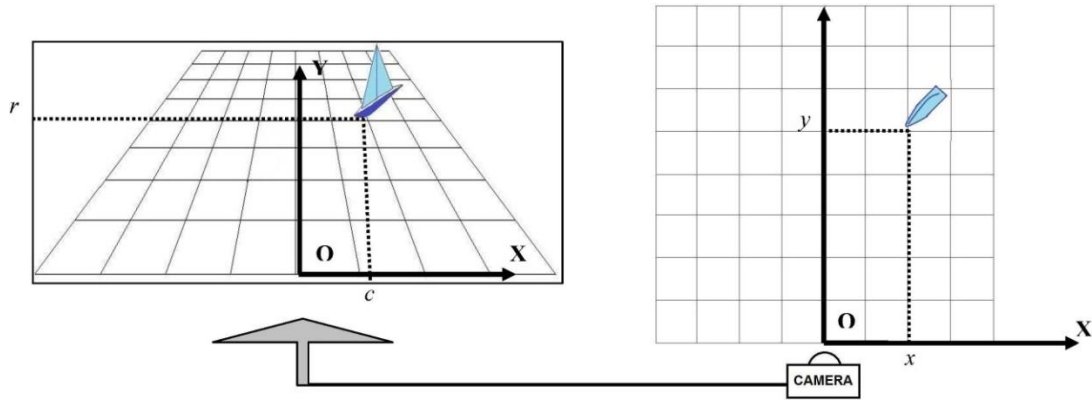
**Figure 3.2:** (a) Camera placed on the shore of the reservoir; (b) An image of the 3D scene as it is seen by the camera.

Although in this work the method is tested with a boat navigating on a water planar surface, it is also valid for positioning different objects in planar surfaces. Based on this fact, we exploit the following two properties derived from the perspective projection, which is the one provided by the optical system attached to the camera:

a) The lowest the object is located in an image the closer it is to the observer. In other words, there is a relation between the row of the image where the object is located and the distance to the observer. This relation is modeled as a mathematical function which links the row of the boat in the image with the real 3D *y-coordinate*. It is to be estimated as explained below.

b) The closer an object is to the observer the wider it seems in the image. That is, horizontal dimensions of an object in an image depend not only on the real dimensions of the object themselves but also on how far the object is to the observer, i.e. a big object which is far away from the observer may seem smaller in the image than a narrow object closer to the observer. This means that the real 3D *x-coordinate* of an object depends on the column but also on the row where the object appears in the image. This second relation is the other mathematical function to be estimated, as explained below.

Considering these two facts, all we have to do is to find these two mathematical relations between the position of an object in an image, identified by its row  $r$ , and column  $c$ , and its real 3D position ( $x$ - $y$  coordinates) in the scene. Part left in *figure 3.3* displays a boat on the image located at a pixel with coordinates  $(r, c)$  in the image. Right side in *figure 3.3* identifies the real 3D  $x$ - $y$  coordinates in the scene.



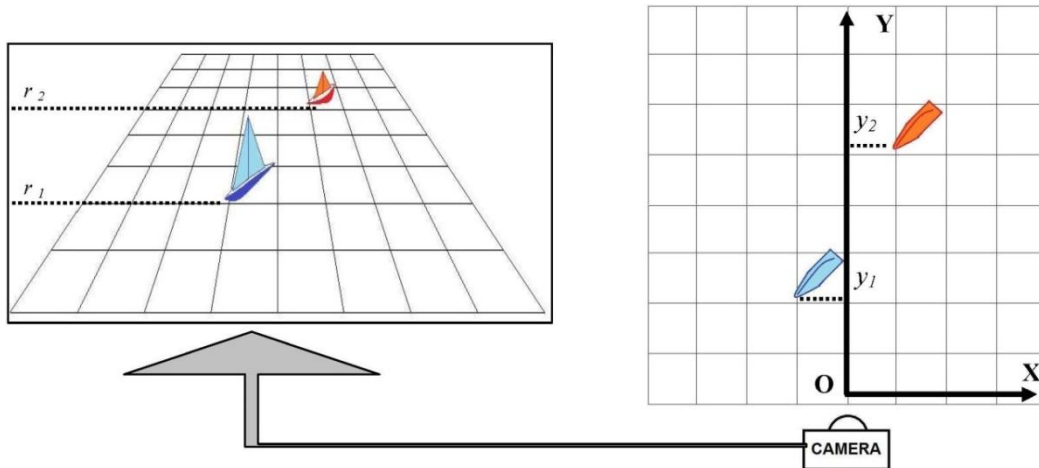
**Figure 3.3:** Left: Image taken by the camera with the row  $r$ , and the column  $c$ , of the boat as it appears in the image. Right: real coordinates of the boat in the scene with its  $x$ - $y$  coordinates with respect the origin  $O$ .

Since  $y$ -coordinate depends only on the row  $r$ , but  $x$ -coordinate depends on the row  $r$ , and also on the column  $c$ , we have to consider two cases separately to find these dependencies: one for the  $Y$ -axis and another for the  $X$ -axis.

a)  $Y$ -axis:

$Y$ -coordinate, which represents distance to the  $X$ -axis, only depends on the row where the pixel is located, that is, the closer the object is to the observer the lower it appears in the image. From *figure 3.4*, we can see that for any two coordinates  $y_1$  and  $y_2$  with their respective rows in the image  $r_1$  and  $r_2$ , if  $y_1 < y_2$  then necessarily  $r_1 < r_2$ . Therefore  $y$ -coordinates depend only on the row where they are placed in the image. The above relative relations can be established under the assumption that objects are placed on a planar surface on the scene.

Now the goal is to find the mathematical function that relates each  $y$ -coordinate with the corresponding row  $r$ . With such purpose, we take different images of the boat in different positions over the planar water surface. These positions are chosen so that they cover the full surface to be analyzed, i.e. with distances from 0 to 150 m from the camera, 150 m is the maximum distance considered in our experiments imposed by the resolution of the camera and the height at which it is placed.



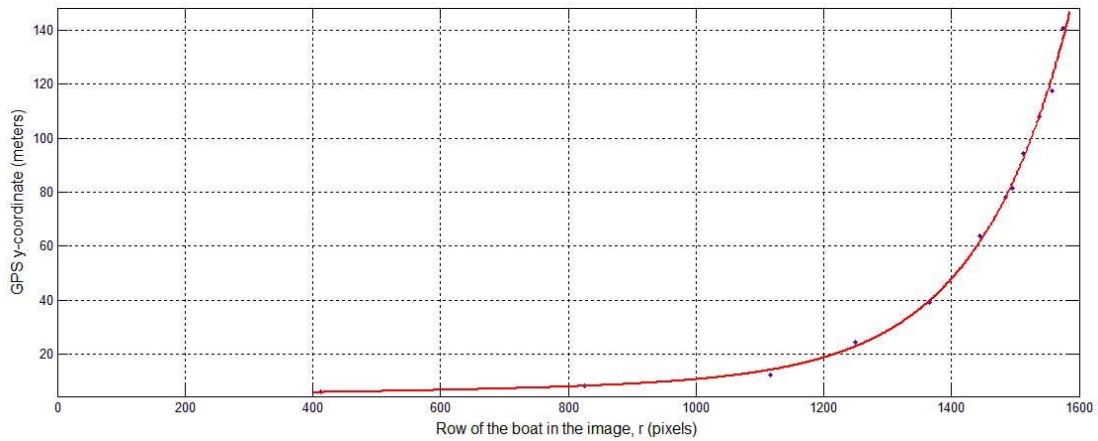
**Figure 3.4:** left: the lowest the object is placed in the image (blue boat, row  $r_1$ ) the closer it is to the observer (right).

The real 3D position of the boat,  $y$ -coordinate, is known thanks to the GPS onboard the boat and the row of the boat in the image is obtained directly from the images by locating a marker point in the boat. Twelve images were acquired for estimating the mathematical functions. *Table 3.1.* displays for each image in *column 1*, positions of the boat at each 3D location, in *column 2* and *column 4*, i.e.  $y$  and  $x$ -coordinates respectively. *Column 3* and *column 5* displays the corresponding image coordinates in pixels. *Column 6* displays what we call ratio ( $R$ ) concept expressed in pixels/meter, which determines the absolute value of the ratio between each column  $c$ , and its associated  $x$ -coordinate.

# Image	$y$ -coordinate (m)	Row in the image, $r$ (pixels)	$x$ -coordinate (m)	column in the image, $c$ (pixels)	Ratio pixels per meter $R=c/x$ (pixel/m)
1	6,3	413	2,8	742	265,0
2	8,2	825	10	1083	108,3
3	12,3	1.116	5	308	61,6
4	24,5	1.249	3,7	146	39,5
5	39,3	1.365	17,5	562	32,1
6	63,6	1.445	9,1	266	29,2
7	78,3	1.484	39,1	1110	28,4
8	81,8	1.496	49,8	1384	27,8
9	94,5	1.513	47,7	1302	27,3
10	108,0	1.538	11	296	26,9
11	117,4	1.558	46,2	1224	26,5
12	140,8	1.574	35,6	935	26,3

**Table 3.1:** Twelve images of the boat in different positions, with the  $y$  and  $x$ -coordinates obtained with the GPS, in meters, for each position and their corresponding rows  $r$ , and columns  $c$ , in pixels over the image. The ratio value  $R$  is also displayed.

Figure 3.5 displays the relation between the  $y$ -coordinate measured with the GPS in meters against the row  $r$  in pixels, i.e.: *column 2* vs. *column 3* from *table 3.1*.



**Figure 3.5:** Relation between  $y$ -coordinate, distance to the  $x$ -axis of the boat, and the row  $r$ , where it is located in the image.

Now, the goal is to find the mathematical function describing the relation between the  $y$ -coordinates and rows  $r$ , expressed as the following model,

$$y = F_y(r) \tag{3.1}$$

a)  $X$ -axis:

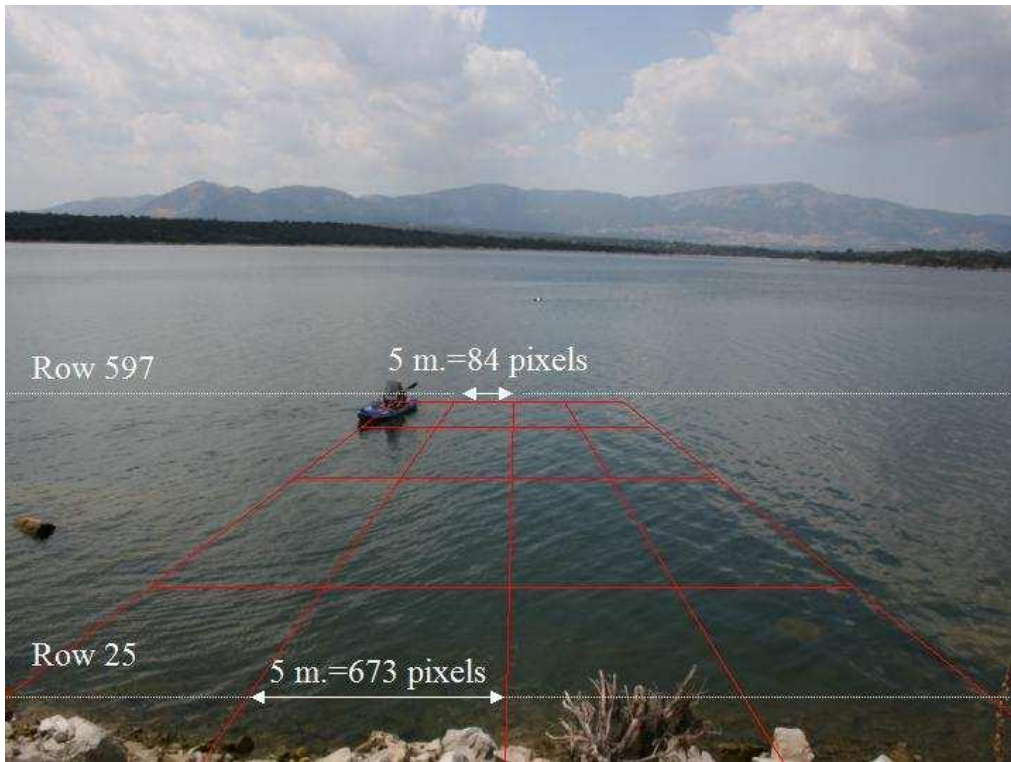
$X$ -coordinate of an object, or equivalently distance to the  $Y$ -axis, as mentioned before, not only depends on the column where the object is located but also on the row. To understand this fact we take a look at *figure 3.6*. that illustrates it with a simple pedagogical example. Indeed, in the horizontal axis ( $X$ -axis) any given distance, for example 5 meters in row 25 is covered with 673 pixels, however the same distance in a higher row (row 597) is covered with 84 pixels.

Therefore the 3D real  $x$ -coordinate of the ship is a function of the column and the row where it appears mapped in the image, which can be expressed as follows,

$$x = F_x(f, c) \tag{3.2}$$

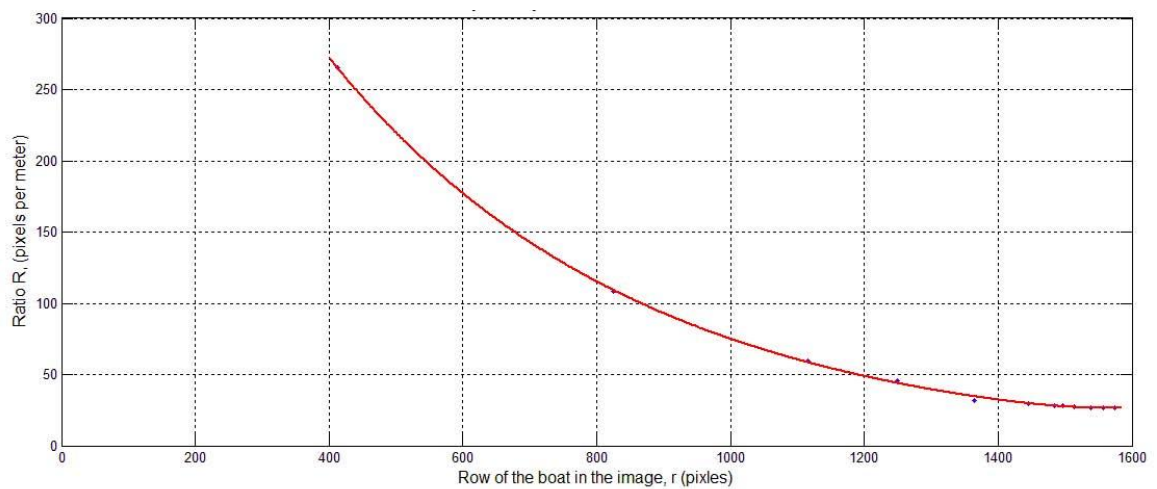
Because of the double dependency of  $x$  from  $c$  and  $r$ , to find how objects width in the  $X$ -axis varies with the row of the object in the image we apply a normalization by computing the ratio  $R$ , of pixels per meter in the  $X$ -axis. Thus, for every image we calculate the ratio as in equation (3.3).

$$R = \frac{c}{x} \quad (3.3)$$



**Figure 3.6:** Horizontal distance depends on the column and also on the row of the image.

Figure 3.7 displays the ratio  $R$  against the row  $r$ , with data in *column 6* and *column 3* respectively in *table 3.1*. Now the goal is to find an analytical function that relates the ratio and the row as follows,  $R = F_R(r)$ .



**Figure 3.7:** Relation between the ratio  $R$  (pixels/meter) and the row of the boat in the image,  $r$ .

From equations (3.2) and (3.3) we can find the  $x$ -coordinate of the boat expressed as a function of  $r$  and  $c$  as expressed in equation (3.4).

$$x = F_x(r, c) = \frac{c}{F_R(r)} \quad (3.4)$$

Once functions  $F_y(r)$  and  $F_x(r, c)$  have been estimated, following the procedure described above, the first stage is finished. Now we can calculate the real  $x$  and  $y$ -coordinates of any pixel in the image by only knowing the row  $r$ , and the column  $c$ , where it is located in the image. This is the second stage mentioned above.

### 3.4. Experimental results

For estimating the above two functions samples have been taken in Valmayor reservoir, located in Madrid (40° 33'00.77'' N, 4°03'07.06''O). The experiment was done in July 2011 with the following equipment used:

- a) Camera Canon EOS 400.
- b) Differential GPS system, Hemisphere GPS A221, A220.
- c) A 4 meters long rowing boat with a marker point used to identify the boat in the images with a GPS system onboard.
- d) Two radio communicators to synchronize GPS readings with the image capture.

The experiment is summarized in the following steps:

- a) A camera is placed on the shore at a certain height from the water surface. The higher it is placed, the more accurate on finding pixels values from the image we will be. This is because if we place it in a high position the water will cover more area in the image than placing it at the water surface level. In our case we placed it on a small hill close to the lake side. Knowing the numeric value of the height is not necessary for calculations.
- b) The boat is moving inside the field of view of the camera.
- c) Different images are taken as the boat moves. For every image acquired, the boat *GPS coordinates* are also stored. In our case, twelve images were taken and used as samples to fit the functions, *table 3.1*.

Applying the method described in *section 3.3*, the following analytical functions, with different parameters, were tested for the fit using the procedure *cftools* from Matlab (*Matworks 2011*): Polynomial, Power, Fourier and Exponential. After different experiments, the best fitting was obtained with the exponential functions given in equations (3.5) and (3.6).

$$y = F_y(r) = 4,603e^{0,000602r} + 0,002255e^{0,006936r} \quad (3.5)$$

$$x = F_x(r, c) \boxminus = \frac{c}{R} = \frac{c}{F_R(r)} = \frac{c}{650,4e^{-0,002173r} + 3,01 \times 10^{-8}e^{0,01209r}} \quad (3.6)$$

Function (3.5) was obtained with the following goodness of fit: SSE: 69.45, R-square: 0.997, Adjusted R-square: 0.9959, RMSE: 2.946. Function (3.6) was obtained with the following goodness of fit: SSE: 34.59, R-square: 0.9993, Adjusted R-square: 0.9991, RMSE: 2.079.

After this estimation, we have tested our approach with 10 additional images captured from the same scene, which are different from the ones used for the fitting above. For testing purposes we compare *y-coordinates* provided by the GPS and the same coordinates calculated with  $F_y(r)$ . *Table 3.2* displays the accuracy of our method with the absolute and relative errors shown in *table 3.2*.

Image number	<i>y-coordinate</i> given by the GPS (m)	<i>y-coordinate</i> estimated with $F_y(f)$ (m)	Absolute error (m)	Relative error (%)
13	6,08	6,16	0,08	1,32
14	7,12	7,06	0,06	0,84
15	7,58	7,44	0,14	1,85
16	9,50	9,70	0,20	2,11
17	17,41	17,83	0,42	2,41
18	22,18	21,61	0,57	2,57
19	67,13	69,48	2,35	3,50
20	82,88	86,28	3,40	4,10
21	105,16	100,01	5,15	4,90
22	133,92	126,96	6,96	5,20
		<b>Mean</b>	<b>1,93</b>	<b>2,88</b>

*Table 3.2: GPS (real) y-coordinates, estimated y-coordinates and their absolute and relative error.*

On the same way, *table 3.3.* shows the values of the  $x$ -coordinates provided by the GPS, the estimated coordinates using  $F_x(c,r)$  and the absolute and relative errors.

Image number	$x$ -coordinate given by the GPS (m)	$x$ -coordinate estimated with $F_x(f,c)$ (m)	Absolute error (m)	Relative error (%)
13	-1.34	-1.32	0.01	1.1
14	-1.69	-1.72	0.02	1.4
15	-2.05	-2.09	0.04	1.9
16	-1.55	-1.58	0.03	2.1
17	0.52	0.51	0.01	2.4
18	1.14	1.10	0.03	2.9
19	12.22	12.55	0.33	2.7
20	24.56	25.32	0.76	3.1
21	33.75	32.64	1.11	3.3
22	44.71	43.41	1.30	2.9
		<b>Mean</b>	<b>0.37</b>	<b>2.38</b>

**Table 3.3:** GPS (real)  $x$ -coordinates, estimated  $x$ -coordinates and their absolute and relative error.

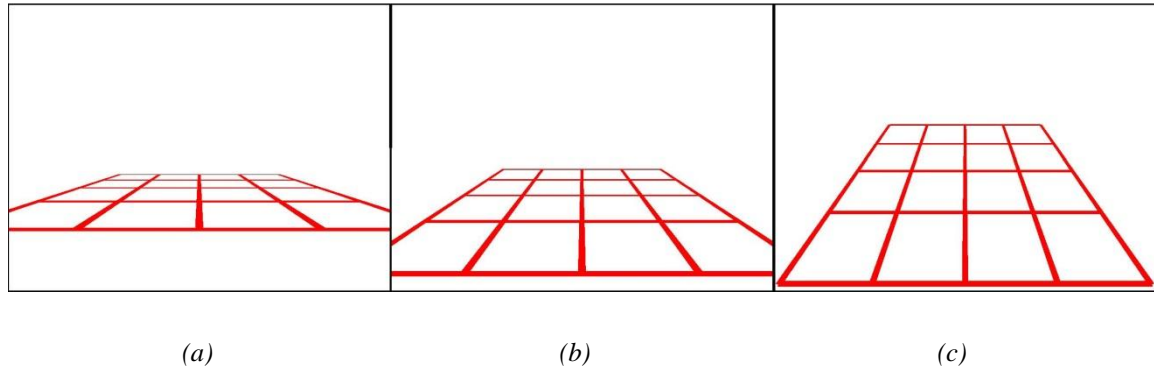
From *tables 3.2* and *3.3* we can see how errors increase as  $y$ -coordinate increase. This is because as we increase distance of the object in the image, the ratio  $R$  decreases which means that we have less pixels resolution for the same distance, as explained in *figure 3.6.*

On the other hand, error may be reduced by increasing the height of the camera and modifying the pitch angle so that our scenario covers as much area in the image as possible (*figure 3.8.*). This is because the more the rows of pixels the scenario covers in the image the more the number of rows to assign to any distance and therefore the better the accuracy.

The following figure is a computer simulation of a four by four meters grid. Camera has been placed at 4 meters away from the grid. Focal length is 10 mm and camera height and pitch angle on every image, from left to right, are: 0,5 m 0 degrees (left), 1 m 6 degrees (center), 2 m 23 degrees (right) respectively.

Additional experiments in different days, i.e. under different conditions, with more than 200 images have provided similar accuracies than the ones displayed in *table 3.3.*

Results shown in this section correspond to the experiment in the reservoir, however this method has been tested with images of objects in long indoor corridors (40 meters long) and small rooms (3 meters long) with better results due to the accuracy in the measures of the samples that are needed to fit the analytical functions and the ratio  $R$ .



**Figure 3.8:** From (a) to (c), different images of the same scenario taken from different heights and pitch angles of the camera. Notice that as height increases we have more pixels per row for the same distance.

### 3.5. Conclusions.

This work has presented an alternative method not only for depth estimation and distance measurement but also for real  $x$ - $y$  coordinates estimation for planar surfaces.

It is achieved with a single camera whose intrinsic and extrinsic parameters are not needed. The only requirement is a previous adjustment of the images to find the analytical functions. Once functions are found the camera must be fixed.

A number of samples is needed to find the functions  $F_y(r)$  and  $F_x(r,c)$ . The number of needed samples will depend on the accuracy of the measures, the resolution of the images and the field of vision we want to cover.

This method, based on a single camera, can be easily used in other applications for locating objects placed on planar surfaces i.e.: sea, flat and crop fields, indoor floors, planar roads among others.

## CAMERA SENSOR ARRANGEMENTS FOR CROP AND WEED DETECTION ACCURACY IN AGRONOMIC IMAGES

### 4.1. Introduction

The use of robotic systems, equipped with vision-based sensors, for site-specific treatments in Precision Agriculture (PA) is seeing continuous growth. A common practice consists of image processing for weed and crop identification. Both, crop lines and weed identification are used for selective treatments (*López-Granados, 2011; Romeo et al., 2013a, 2013b; Guerrero et al., 2012c; Gée et al., 2008; Zheng et al., 2009; Montalvo et al., 2012a; Guijarro et al., 2011; Burgos-Artizzu et al., 2009; Sainz-Costa et al., 2011; Tellaeche et al., 2008; Jones et al., 2009a,b; Tellaeche et al., 2008*) additionally crop line identification is also used for tractor guidance (*Li et al., 2009; Reid and Searcy, 1997; Billingsley and Schoenfisch, 1995, 1997; Rovira-Mas et al. 2003*). Crop line and weed detection is an important issue related with the application of machine vision methods in agriculture, and consequently has attracted numerous studies in the area.

Ratios of greenness to soil determine what is known as density. The goal is always to detect crop/weed densities and also crop lines from the imaged spectral components. These tasks must always be carried out using a given camera-based sensor with the highest accuracy and robustness possible and under adverse and highly variable outdoor illumination conditions, which are the natural environmental conditions in agricultural fields.

Accuracy may be critical for determining whether specific treatments are required for weed control or for vehicle guidance when required. Accuracy in this matter is directly

related with intrinsic and extrinsic camera parameters (*Fu et al. 1988*), and also with some factors caused by the uncontrolled outdoor illumination (*Jeon et al. 2011*).

The camera-based sensor consists of three main parts, namely: (a) CCD device; (b) optical lens and (c) ultraviolet and infrared cut filters for controlling the input of only those wavelengths of interest. Some camera parameters are fixed for the system's requirements according to the goals of the application. Sometimes this leads to the choice of a specific sensor, as described below in *section 4.3.1*, with their intrinsic parameters already predefined and perhaps some extrinsic ones too. This work is concerned with the definition and analysis of those unfixed extrinsic parameters and also with factors to control the adverse effect of the illumination, always with accuracy purposes for greenness identification and later crop line and weed detection.

Regarding the extrinsic parameters, they depend directly on the physical vision system, which is installed on-board the tractor pointing to a selected area in the field, ahead of the tractor. This is an important issue in PA (Precision Agriculture), which is addressed in this chapter and described in *section 4.3.1*.

Regarding the illumination factors, we can point out that the processing of agronomic images becomes a difficult task because they are always captured under uncontrolled illumination conditions (*Jeon et al. 2011*). Indeed, in outdoor environments a variety of weather conditions may appear, i.e., highly sunny or cloudy days with different intensities, clear days alternating with different cloud densities, etc.

Moreover, it is well known that in outdoor environments, particularly in sunny days, infrared radiation enters the sensor impacting the different spectral channels. The control of illumination factors is addressed in *section 4.3.2*.

Based on a given camera-based sensor with its corresponding accessories, we study the image accuracy for crop line and weed detection in agronomical images with specific reference to maize crops, where an area 3 m wide must be covered. This accuracy is studied from two points of view, making the main findings of this work: (a) geometrical arrangement, based on extrinsic parameters and (b) software corrections for improving the image quality, derived from the uncontrolled illumination in this kind of outdoor environments.

The results of this work were published in *Romeo et al. (2013a)*, *Guerrero et al. (2012c)* and *Montalvo et al. (2011)*.

## 4.2. Materials

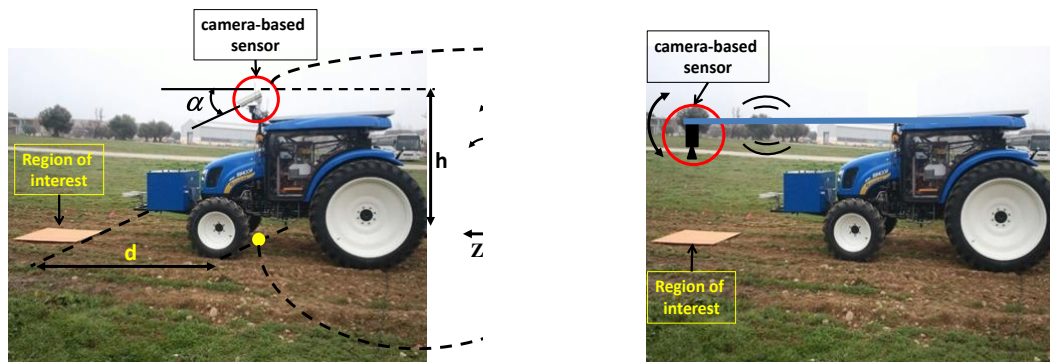
The camera-based sensor consists of three essential physical parts: (a) CCD-based device embedded in a housing with its electronic equipment and interfaces for power supply and to the computer; (b) optical lens and (c) ultraviolet and infrared cut filter. *Figure 4.1* displays these parts assembled as a whole in *figure 4.1(a)* and separated in *figure 4.1(b)*.



**Figure 4.1:** CCD sensor, lens and UV/IR cut filter: (a) integrated; (b) separated.

The CCD is a Kodak KAI 04050M/C sensor with a Bayer color filter with GR pattern; resolution of  $2,336 \times 1,752$  pixels and  $5,5 \times 5,5 \mu\text{m}$  pixel-size. This device is part of the SVS4050CFLGEA model (SVS-VISTEK 2014) which is robust enough and very suitable for agricultural applications. This device offers several externally controlled possibilities: (a) exposure time, which determines the time taken to capture the image; (b) Red, Green and Blue gains, where a value can be set for each channel, including gains auto-calculation; (c) definition of specific Regions Of Interest (ROIs); (d) information about the operating temperature. This Gigabit Ethernet device connected to a cRIO-9082 with dual-core controller, 1.33 GHz and LX150 FPGA running under LabView (*National Instruments 2014*) is robust enough and specifically designed for real-time processing, so both features are very suitable for our agricultural application. Because the application occurs in harsh environments (containing dust, drops of liquid from sprayers, *etc.*) it is encapsulated in a housing with IP65 protection and internally

equipped with an automatic fan which is triggered if the temperature surpasses 50 °C; this housing is displayed in *figure 4.2(a)* indicated by the label camera-based sensor.



**Figure 4.2:** Camera-based sensor arrangement with a ROI in front of the tractor: (a) near the mass center of the tractor with referential coordinate systems; (b) Zenithal position.

The optical system consists of a Schneider Cinegon 1.9/10-0901 lens (*Schneider Kreuznach 2014a*), with manual iris aperture (f-stop) ranging from 1.9 to 16 and manual lockable focus, providing high stability in the agricultural environment. It is valid for sensor formats up to a diagonal value of 1", i.e., maximum image circle of 16 mm, and is equipped with an F-mount which can be adapted to C-mount. The focal length is fixed at 10 mm. Its field of view is above 50° with object image distance from infinity to 7.5 mm, which allows the mapping of a width of 3 m as required for our application. Its spectral range varies from 400 to 1,000 nm, i.e., visible and near-infrared (NIR). Under this optical system the images are captured with perspective projection (*Fu et al. 1988*).

As mentioned before, our system works in adverse outdoor agricultural environments where the natural illumination contains a high infra-red component. The sensor is highly sensitive to NIR radiation and to a lesser extent to ultra-violet (UV) radiation. The NIR heavily contaminates the three spectral channels (Red, Green and Blue) producing images with hot colors. This makes identification of crop lines and weeds unfeasible because during the treatments these structures are basically green. To avoid this undesired effect, the system is equipped with a Schneider UV/IR 486 cut-off filter (*Schneider Kreuznach 2014b*). Its operating curve specifies that wavelengths below 370 nm and above 760 nm are blocked, i.e., both UV and NIR radiation. Despite this blocking effect, a *vignetting* effect remains, requiring correction as described below.

More than 2,000 images have been acquired at the Institute of Agricultural Sciences and the CAR facilities in Arganda del Rey (Madrid) and in the School of Agronomic Engineering in Madrid, all on different dates, during April, May, June and July in years 2011, 2012 and 2013 from maize fields and the last ones on November 2012 and January 2013. No maize crops are available at this time of the year. Because our application is specifically designed for maize crops, crop lines have been made by mowing six 80 meter long lines among weeds. Lines are separated 75 cm from each other like in real maize crops.

To quantify the number of pixels with the maximum accuracy as possible, a bright orange colored cardboard of  $1 \times 1 \text{ m}^2$  is used. This cardboard defines the physical ROI to be imaged with a peculiar color, which is not present in agronomic images. It is placed in front of the tractor at different distances. As mentioned before, these distances define one of the extrinsic parameters involved in this study related to accuracy from a geometric point of view.

## 4.3. Methods

### 4.3.1. Accuracy from the point of view of extrinsic parameters

As mentioned before, the vision system is installed on-board a tractor pointing to a selected area ahead. *Figure 4.2(a)* displays the camera configuration in the tractor, specifically devoted to treatments in maize fields under the project funded under the Seventh Framework Program: *RHEA (2014)*. In this project an important issue is related to the area to be treated, because the implement used for applying treatment, covers four crop lines spaced 0.75 m each, so the total wide area to be captured for the camera must be at least 3 m in width. Although it is always possible to combine different camera resolutions with different focal lengths, an acceptable solution for positioning the camera, at the same time the width of the scene is captured fulfilling this requirement, is the one displayed in *figure 4.2(a)*.

Under this configuration, the origin of the world coordinate system *OXYZ* is located exactly in the ground with its axes oriented as displayed. At a height *h* from *O* is placed the origin *O* of the coordinate system *oxyz* attached to the plane of the image, i.e., coinciding physically with the CCD. Angles pitch  $\alpha$ , roll  $\theta$ , and yaw  $\beta$ , define the three degrees of freedom of the image plane with respect to the *OXYZ* system. They

define three extrinsic parameters, which are the basis for building the rotation matrix. A fourth extrinsic parameter is  $h$ , which defines the height of the camera; it is involved in the translation matrix. Although the position of the camera is fixed under this tractor configuration, some displacements along the vertical are still possible, so that  $h$  can be considered variable.

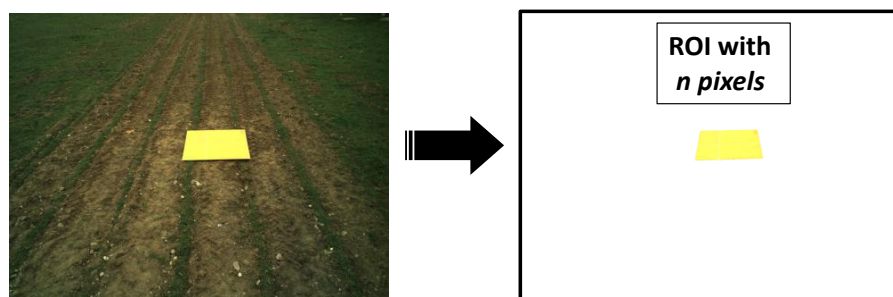
The tractor moves along the crop lines and it processes in real time the images determining the percentages of weeds and the crop line location. Tractor speed is a critical factor since the faster the tractor moves along the crop lines, the faster the algorithms for detection have to be or else the further away the selected area to be focused has to be, with the resulting unavoidable loss of accuracy.

Ideally the best pitch angle would be  $90^\circ$ , i.e., with camera pointing downwards (zenith orientation). This camera arrangement has been used for discrimination between crops and weeds in sugar fields (*Astrand and Baerveldt 2002*) and its performance has been reported in (*Baret et al. 2010*). In our application, this pitch angle arrangement would imply that the camera would have to be placed under the tractor. This solution requires a lens which allows a broad field of view to cover the required width of 3 m and that images processed on real time would have to be processed nearly instantaneously. Another solution would be placing the camera fixed on a bar (*figure 4.1(b)*) at a certain distance ahead of the tractor so that algorithms would have more time to process the images as the tractor reaches the processed area. But this performance would produce high oscillations with the tractor's movement which would result in useless blurry images. Therefore, an acceptable position for the camera on the tractor is the one as close to its mass center as possible to avoid oscillations and with the pitch angle adjusted so that the camera points to the ROI with the highest possible accuracy.

In the same way, the shorter the distance of the region of interest to the tractor, i.e., to the origin of the OXYZ, the higher the accuracy obtained in the acquired images. However, the area must not be too close because of the following two constraints: (a) the box in front containing energy accumulators requires a minimum pitch angle to cover the ROI; (b) in the time for image processing; the minimum distance is limited since a very short distance to the tractor would not give time enough for image processing.

Regarding the extrinsic roll  $\theta$ , and yaw  $\beta$ , parameters, we assume they are fixed to zero values because of the plane terrain (roll) and correct guidance (yaw). Obviously, they could be considered to be non-zero if the above constraints are not met. Thus, assuming both extrinsic parameters are known, the image accuracy, based on the above mentioned constraints, depends on the pitch angle  $\alpha$ , the height of the camera  $h$ , and the distance of the ROI  $d$ , so this work studies the image accuracy based on these three parameters, i.e.:  $\alpha$ ,  $h$  and  $d$ , which are related to each other. The goal is to find out the relation among them to determine the camera arrangement needed to achieve the best image accuracy. This represents an extension with regard the study in (Guerrero J.M. et al. 2012), where only the pitch angle was the subject of study, also for image accuracy, under a simulated scenario through Webots (*Cyberbotics. Webots: Robot Simulator 2014*), instead of a real one as now. The use of simulated scenarios is a common practice before using the real ones (*Burgos-Artizzu et al. 2009*). Depending on the agronomic application some systems set beforehand some extrinsic parameters because the system performance is guaranteed, either for detecting weeds and crops (*Jiang et al., 2010; Hague et al., 2006; Bossu et al., 2009*) or for row following (*Astrand and Baerveldt, 2005; Jinlin and Liming, 2010*), including stereovision systems (*Kise et al., 2005; Rovira-Mas, 2011*).

For quantifying the ROI resolution of every image captured under different values of the extrinsic parameters to be studied, this image is segmented and the number of pixels of the orange cardboard is computed based on a segmentation process developed under LabView (*National Instruments 2014*). This process is easy, it is based on the color spectral properties displayed by the peculiar orange color mapped on the image from the original cardboard, every non bright orange pixel is removed, so that the remaining ones in the image belong to the ROI, *figure 4.3*. The higher the number of remaining orange pixels in the image the higher the resolution of the ROI. This is the criterion for measuring the accuracy based on the extrinsic parameters.



**Figure 4.3:** ROI in the field defined by the orange cardboard and number of its pixels.

A set of 125 images have been acquired consisting of five different pitch angles, five different heights and five different distances, all combined with each other. This means that every image corresponds to a different pitch angle, a different height of the sensor camera from the ground and a different distance of the ROI from the sensor camera.

In order to achieve this set of images we proceed as follows:

1. The camera-based sensor is placed in the same position as the one it has in the tractor pointing at the crop lines.
2. The cardboard is placed in the middle of the crop lines at the distance where weeds will be detected once the tractor is in movement (*figure 4.2*).
3. A set of different images is taken varying the pitch angle, the height of the camera from the ground and the distance of the camera to the ROI alternatively.

That is:

```

for  $h = h_1$  to  $h_5$  do
  for  $d = d_1$  to  $d_5$  do
    for  $\alpha = \alpha_1$  to  $\alpha_5$  do
      • Capture image  $I_{hda}$ 
      • Identify the cardboard; apply color image segmentation based on
        the RGB spectral components
      • Compute the area in terms of the number of pixels belonging to
        the identified cardboard.
    end;
  end;
end;

```

### 4.3.2. Accuracy from the point of view of illumination factors

As already mentioned, the quality of the images is highly dependent on illumination, which affects highly the segmentation algorithms and indirectly the image accuracy from the point of view of crop line detection and weed identification. Thus, in addition to the extrinsic parameters, the image accuracy can be studied from the point of view of image quality. In this work, we make the following contributions to this topic:

Controlling the amount of light impacting the sensor. The diaphragm or iris aperture is fixed, because no auto-iris lens is available in our system, and the amount of

light is controlled via the exposure time, which is a facility provided by the camera sensor (*Montalvo y col. 2013b*).

Applying a software correction procedure for minimizing the phenomenon known as lens *vignetting* effect, where the brightness decreases at the periphery in the image compared to the center of the image. This effect can be produced by different causes, including optical properties of the lens itself, particularly in lenses with wide aperture. When we apply a cutting infrared filter, part of the red wavelengths are attenuated causing an excess of greenness, particularly where *vignetting* is more evident, i.e., at the periphery of the image and more specifically on the four corners.

#### 4.3.2.1. Exposure Time

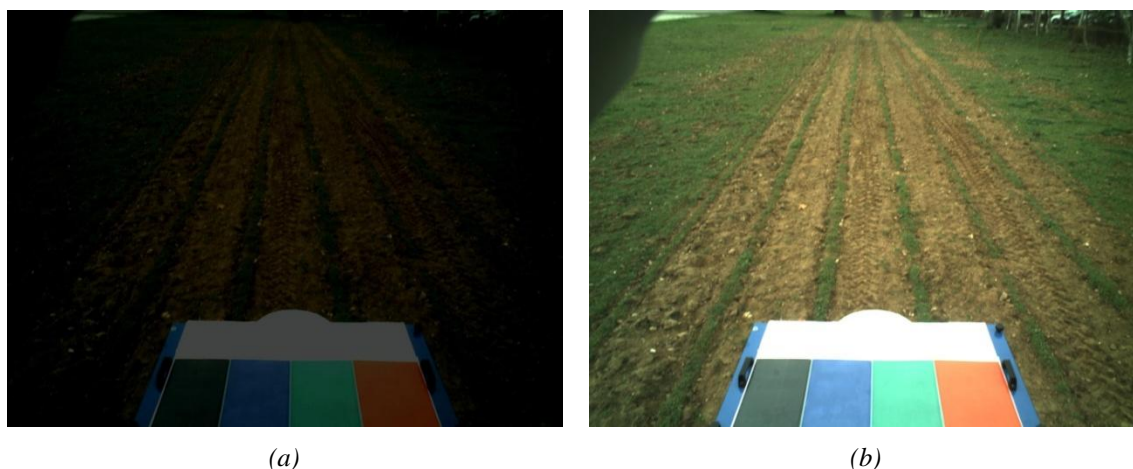
A panel with four colors is placed in front of the tractor and inside the field of view of the camera. At this stage we assume known the extrinsic parameters for the best accuracy, i.e., geometric camera arrangement with  $\alpha$  and  $h$  values. Thus, we know the position of the panel in the image and also the distribution of colors on it (*figure 4.4*). The white part of this panel serves as reference for controlling the exposure time.

The procedure consists of the following steps:

1. Capture an image.
2. Perform a sampling inside the white part and compute the average value for the three spectral RGB components, i.e.,  $R$ ,  $G$ , and  $B$ .
3. Adjust the Exposure time  $E_t$ , so that the highest of the three maximum average values,  $H = \max R, G, B$ , falls on a given interval to ensure that the sensor is sufficiently excited but without reaching saturation. In our sensor with 8 bits in resolution per channel, the maximum value is  $M = 2^8 = 255$ , hence the interval is as follows:  $aM \leq H < bM$ , where  $a$  and  $b$  define the lower and upper limits, in our experiments we have verified that  $a = 0.90$  and  $b = 0.98$  suffices.
4.  $E_t$ , which represents the previous exposure time, is adjusted as follows: if  $H < 0.90M$  then  $E_t = (1 + p)E_t$ ; if  $H > 0.98M$   $E_t = (1 - p)E_t$  otherwise  $E_t$  does not need updating; where  $0 < p < 1$  represents the fraction of adjusting, set to 0.20 in our experiments.

5. If after the adjustment  $H$  does not fall inside the interval specified, a new image is captured with the last  $E_t$  updated and *steps 1 to 4* are repeated again.

*Figure 4.4(a)* displays an image captured with an  $E_t = 2 \times 10^4 \mu\text{s}$  where the white panel displays values requiring  $E_t$  adjustment; *figure 4.4(b)* shows the same image acquired with  $E_t = 3.5 \times 10^4 \mu\text{s}$ . We can see how the whole image, and particularly the white panel, has increased their values resulting in a brighter image.



**Figure 4.4:** Images captured from the same scene with two different exposure times: (a)  $E_t = 2 \times 10^4 \mu\text{s}$ ; (b)  $E_t = 3,5 \times 10^4 \mu\text{s}$ .

#### 4.3.2.2. Vignetting correction.

As specified by the manufacturer, the Schneider UV/IR 486 cut-off filter (Schneider Kreuznach 2014b) is based on what is known as thin-film technology containing more than thirty coats on one of its sides and a multi-resistant coating on the opposite one. The incidence angle of these rays in the periphery of the filter is greater than in the center and they must travel longer distances along the different layers of interference. This effect is more pronounced the shorter the focal length of the lens, *i.e.*, lenses with wide-angles. This occurs in our case with a lens of 10 mm.

The effect is that the real color is displaced toward green and blue at the expense of the red one. Thus, in order to correct this effect we have designed an image gray pattern ( $P$ ) with values ranging in  $[0, 1]$ , where at the center of the pattern ( $c_x, c_y$ ) the value is zero and achieves its maximum value of one at the four corners. The size of this

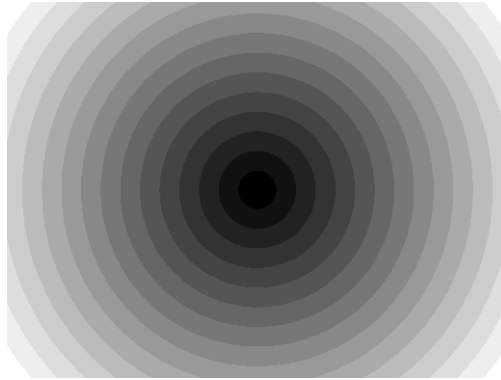
pattern is exactly the one of the image and the value of each pixel  $i$  located at  $(x, y)$  is computed as follows:

$$d(x, y) = \sqrt{(x - c_x)^2 + (y - c_y)^2} \quad (4.1)$$

*Figure 4.5.* displays this pattern. Given a channel  $R$ ,  $G$  and  $B$  we apply the following operation to obtain the corrected values,  $R_c$ ,  $G_c$  and  $B_c$ :

$$R_c = 1 + K_r P \times R; \quad G_c = 1 + K_g P \times G; \quad B_c = (1 + K_b P) \times B \quad (4.2)$$

where  $\times$  denotes pixel-by-pixel multiplication instead matrix product;  $K_r$ ,  $K_g$  and  $K_b$  represent the trade-off between corrections, based on the behavior of the Schneider UV/IR 486 cut-off filter (*Schneider Kreuznach 2014a*) we have verified, with high level of satisfaction that the following values are appropriate in our experiments, obtained by trial and error:  $K_r = 0.3$ ;  $K_g = 0.0$  and  $K_b = 0.0$ . This means that only the red channel is corrected in our application where crop line and weed detection with high accuracy is the goal



*Figure 4.5: Image pattern used for vignetting correction.*

## 4.4. Results and discussion

This study is concerned with the accuracy of weed and crop line detection in maize fields under the RHEA project (*RHEA 2014*). It has been focused on the definition of three critical extrinsic parameters and also on how to control adverse effects produced by the illumination.

#### 4.4.1. Analysis of extrinsic parameters

*Table 4.1* shows the number of pixels ( $n$ ) obtained for the ROI after applying the method described in *section 4.3.1*, for the set of 125 images. Each image  $I_{hd\alpha}$  is captured under different values for the pitch angle ( $\alpha$ ), height of the camera ( $h$ ) and distance to the ROI ( $d$ ).

Results in *table 4.1* are graphically displayed in *figure 4.6* for convenience and better visibility, with the 125 values obtained for  $n$ . As we can see from the graph, there are five sets of five values each (marked with a red circle) that correspond to the maximum values of  $n$ . All these values were obtained at the distance of 3 meters which leads us to the first conclusion of our study (a quite intuitive but perhaps not obvious one): “*The closer the ROI is from our camera the higher the resolution and therefore the accuracy on weeds and crop lines detection*”.

In order to assess the real results in *table 4.1* against theoretical results, we have developed a simulation program in Matlab R2009b (*The Mathworks 2014*). With such purpose in mind, we define a ROI with identical size *i.e.*,  $1 \times 1 \text{ m}^2$ . The coordinate systems are defined like in the real situation and the focal length is also fixed to 10 mm. The CCD-device specifications are the ones corresponding to the Kodak KAI Sensor provided in *section 4.2*. The free parameters provided to the program are obviously  $h$ ,  $d$  and  $\alpha$ . For each combination of these free parameters we map the four corners from the ROI into the corresponding pixels in the image, obtaining a trapeze. We compute the area of this trapeze and compute the percentage against the number of pixels ( $n$ ) displayed in *table 4.1*. The average value for all combinations of the free parameters in *table 4.1* is 0.0088% with standard deviation of  $7,9 \times 10^{-4}$ . As we can see, simulated and real values are very close, verifying the validity of the real experiments for the proposed sensor arrangement and the configuration of the extrinsic parameters.

$\alpha = 10^\circ$			$\alpha = 20^\circ$			$\alpha = 30^\circ$			$\alpha = 40^\circ$			$\alpha = 50^\circ$		
$d$ (m)	$h$ (cm)	$n$	$d$ (m)	$h$ (cm)	$n$	$d$ (m)	$h$ (cm)	$n$	$d$ (m)	$h$ (cm)	$n$	$d$ (m)	$h$ (cm)	$n$
3.0	210	120,456	3.0	210	119,872	3.0	210	114,152	3.0	210	118,338	3.0	210	121,236
	215	122,547		215	121,622		215	115,351		215	118,645		215	123,568
	220	124,754		220	123,682		220	116,321		220	119,078		220	125,698
	225	126,354		225	125,462		225	117,244		225	119,176		225	127,425
	230	128,452		230	127,442		230	118,089		230	119,357		230	129,423
3.5	210	101,239	3.5	210	98,589	3.5	210	94,523	3.5	210	94,021	3.5	210	94,865
	215	102,569		215	99,745		215	95,325		215	95,685		215	95,862
	220	103,698		220	100,695		220	95,986		220	96,852		220	96,899
	225	103,967		225	100,885		225	96,854		225	97,246		225	97,585
	230	104,693		230	101,210		230	98,752		230	98,563		230	98,865
4.0	210	71,895	4.0	210	70,912	4.0	210	68,108	4.0	210	74,594	4.0	210	74,625
	215	72,698		215	71,992		215	68,812		215	74,729		215	74,987
	220	74,147		220	73,245		220	69,543		220	74,947		220	75,125
	225	74,236		225	73,856		225	69,332		225	74,752		225	75,012
	230	75,100		230	74,342		230	70,811		230	75,219		230	75,652
4.5	210	55,326	4.5	210	54,856	4.5	210	53,865	4.5	210	52,023	4.5	210	51,745
	215	56,314		215	55,123		215	54,169		215	52,869		215	51,658
	220	57,259		220	56,985		220	54,896		220	53,695		220	52,896
	225	58,965		225	59,001		225	55,896		225	54,754		225	54,585
	230	59,996		230	59,896		230	56,987		230	55,625		230	55,632
5.0	210	37,989	5.0	210	36,552	5.0	210	35,648	5.0	210	39,602	5.0	210	39,874
	215	38,654		215	37,472		215	36,182		215	40,226		215	40,325
	220	39,541		220	38,442		220	36,762		220	40,713		220	41,256
	225	39,987		225	39,152		225	37,421		225	41,160		225	42,015
	230	41,253		230	40,232		230	37,895		230	41,581		230	42,569

**Table 4.1:** Number of pixels ( $n$ ) in the ROI obtained for different combinations in the pitch angle ( $\alpha$ ), height of the camera ( $h$ ) and distance to the ROI ( $d$ ).

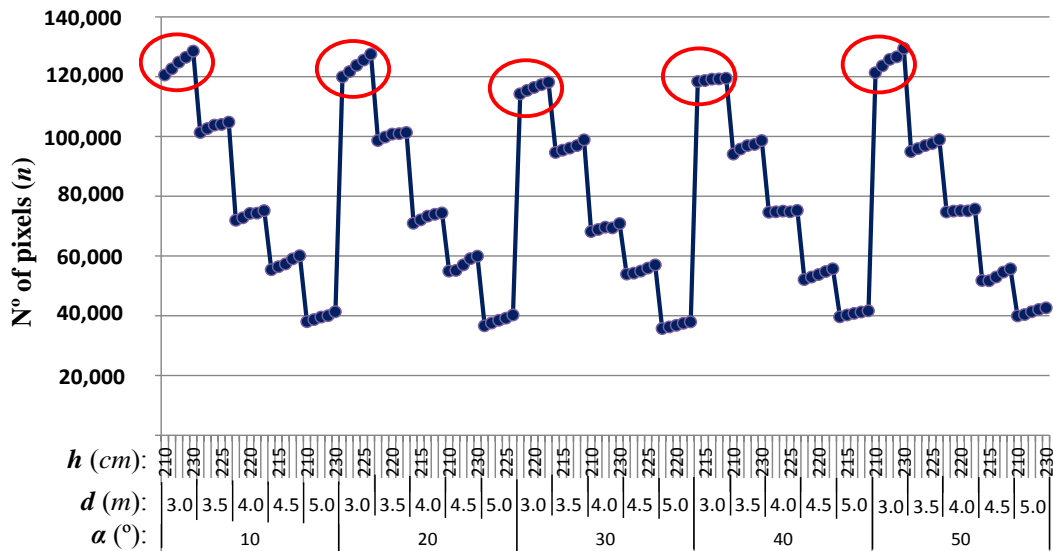


Figure 4.6: Dependency of the number of pixels ( $n$ ) against  $h$ ,  $d$  and  $\alpha$ .

Once we know that distance is a critical parameter for accuracy, we fix the distance of the ROI at 3 m and see how  $n$  varies with the other two parameters. Fixing distances away from the tractor we graph the different values of  $n$  depending on the pitch angle  $\alpha$ , and the height  $h$ , of the camera (figure 4.7).

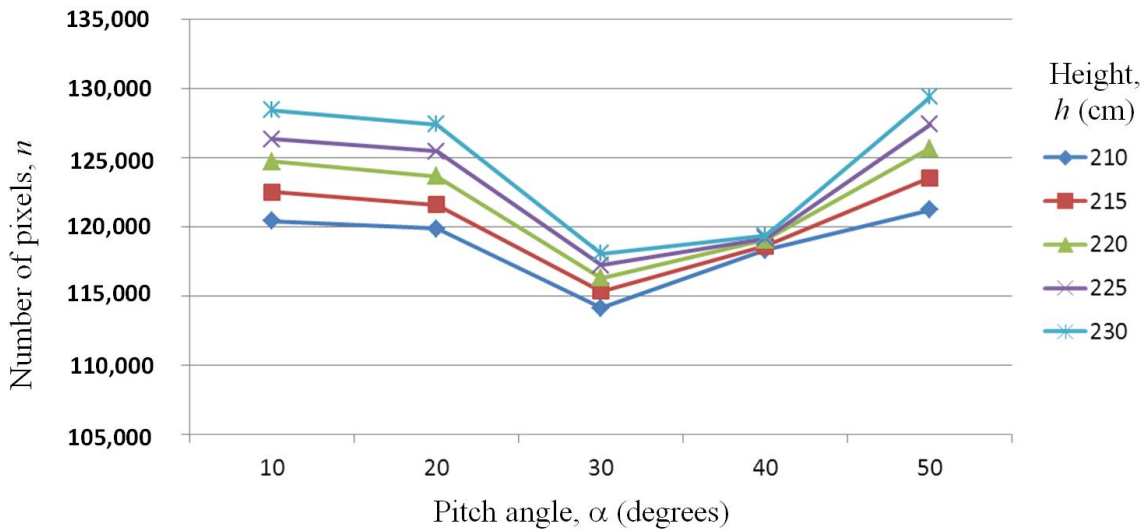
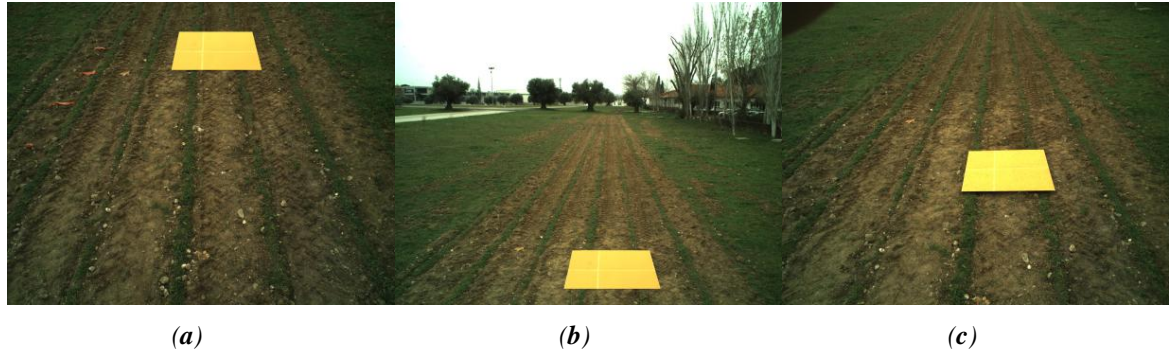


Figure 4.7: Number of pixels  $n$ , against  $\alpha$ , for the five different heights  $h$ , at a fixed distance  $d = 3$  m.

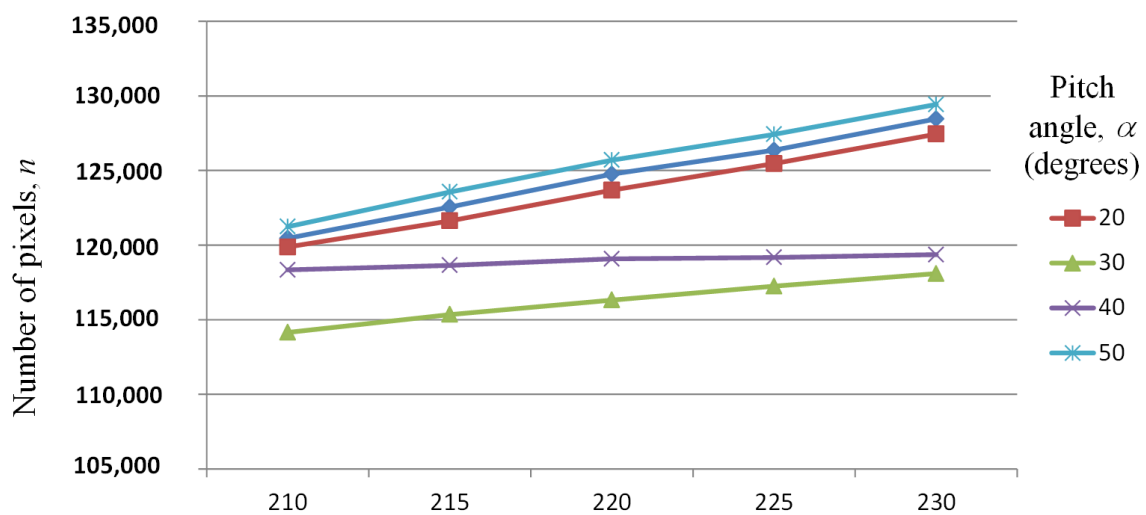
From figure 4.7 we can see that the best pitch angles, in terms of accuracy, are those in which the ROI is placed either at the top of the image which corresponds with  $\alpha = 50^\circ$  (figure 4.8(a)), or at the bottom of the image with  $\alpha = 10^\circ$  (figure 4.8(b)), and the worst pitch angle is that in which the ROI is placed at the center of the image with  $\alpha = 30$  (figure 4.8(c)). The images in figure 4.8. were captured at a height  $h = 215$  cm and a

distance  $d = 3$  m. The above conclusion was already reported in (Reid J.F. and Searcy S.W. 1997), although only considering the pitch angle  $\alpha$ , and without the intervention of the height  $h$ . Thus under this more exhaustive study we arrive at the same conclusion with respect the pitch angle.



**Figure 4.8:** Different images of the ROI (bright orange cardboard) captured with  $h = 215$  cm and  $d = 3$  m under three different pitch angles: (a)  $\alpha = 50^\circ$ ; (b)  $\alpha = 10^\circ$  and (c)  $\alpha = 30^\circ$ .

Regarding the height ( $h$ ) in figure 4.7., we can see that the greatest number of pixels is always obtained with  $h = 230$  cm. This appears a surprising result because without exhaustive analysis one might feel that the lower height objects appear with greater resolution. To clarify this, we again display the number of pixels  $n$ , against the pitch angle  $\alpha$ , and the height  $h$ , but this time we place on the X-axis the height (figure 4.9).



**Figure 4.9:** Number of pixels  $n$ , against  $h$ , for five different pitch angles  $\alpha$ , at a fixed distance  $d = 3$  m. (dot means Decimal point or thousand, please change to comma if dot refer to thousand, the same to figure 4.13)

Indeed, we can see that the accuracy increases as we increase both the height of the camera in the tractor and the pitch angle. This is due to the geometric camera arrangement. *Figure 4.10.* displays a pedagogical example where this is clarified. If we would place the camera at a very close height from the ground, at a fixed distance  $d$  with a given pitch angle  $\alpha$ , we would get an image like the one displayed in *figure 4.10(a)*, however as  $h$  and  $\alpha$  increase with  $d$  remaining fixed, due to a new perspective arrangement, the area of the cardboard that can be seen from the camera also increases, *figure 4.10(b)*.



**Figure 4.10:** Images of the same ROI at a fixed distance and the same pitch angle: (a) with the camera almost at ground level; (b) image obtained at a greater height from the ground.

To emphasize the importance of choosing the correct parameters in order to get the highest possible accuracy for a specific area of the image, we take a look of the maximum number of pixels,  $n_{max}$  and the minimum,  $n_{min}$ , obtained in our study:

$$n_{max} = 129,423 \text{ pixels, } (\alpha = 50^\circ, h = 230 \text{ cm, } d = 3 \text{ m}).$$

$$n_{min} = 35,648 \text{ pixels, } (\alpha = 30^\circ, h = 210 \text{ cm, } d = 5 \text{ m}).$$

This means that the same area of a region in the 3D scene can be imaged into a ROI with a very different number of pixels depending on the parameters involved in this study, *i.e.*,  $\alpha$ ,  $h$  and  $d$ . This can be critical in some agricultural applications. For instance, consider a combination of parameters giving  $n = 129,423$  and a second arrangement with  $n = 35,648$ , this results in a loss of resolution up to 72.5%, *i.e.*, the accuracy reduction.

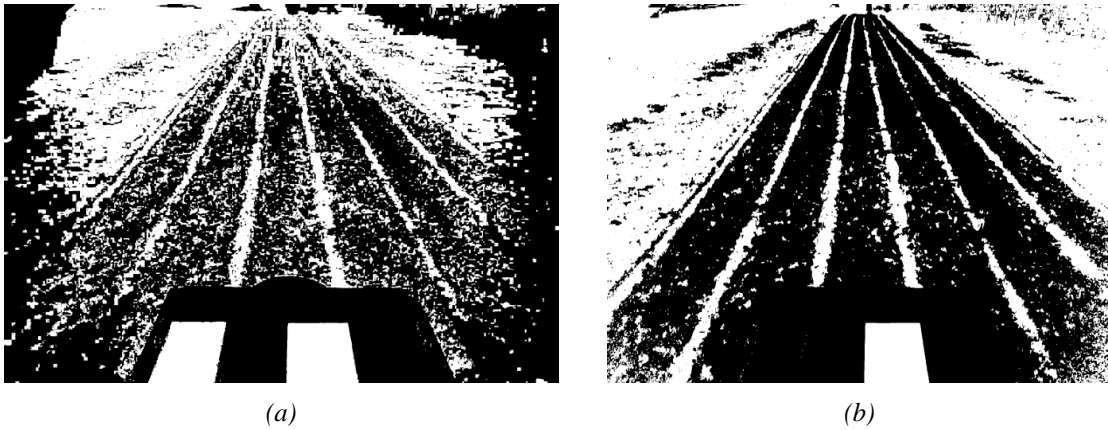
We have studied the best arrangement for the extrinsic parameters and we have found a trade-off between them. Considering that the tractor is in movement at a speed of approximately 4 m/s and our ROI is 5 m deep by 4.5 m wide a good choice of

parameters is:  $\alpha = 20^\circ$ ,  $h = 220$  cm and  $d = 4$  m. The parameters we have chosen give us, according to our study, a resolution of  $n = 56,985$ . If we compare this resolution with the maximum resolution obtained in our study ( $n = 129,423$ ) we realize that, due to constraints not related with the sensor camera, we are giving up 56% of the accuracy. However, had we not considered the importance of these parameters, a worse election of these would have led us to a reduction of 72.5% of the resolution.

So, apart from the intrinsic parameters there are other parameters to be considered with a critical influence on the accuracy of the ROI. This means that when the sensor specification is provided we need to determine, depending on the application, which is the best arrangement of the extrinsic parameters. In this study we have only considered the pitch angle, but also accuracy depends on the yaw and roll angles. These angles should be considered if the camera arrangement involves its consideration or when the tractor must work in a sloping ground.

#### **4.4.2. Analysis of illumination factors**

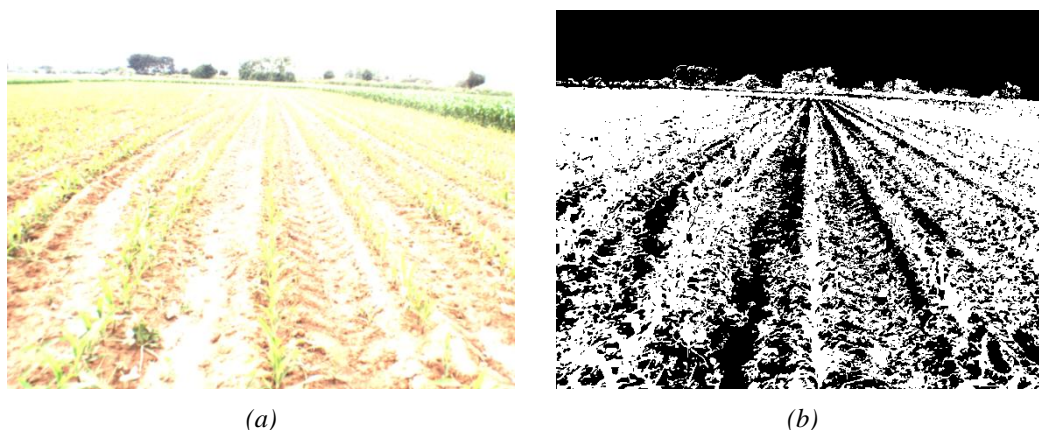
The second group of factors affecting the image accuracy, now in terms of the quality of the image based on the RGB spectral components, is the one derived from the illumination factors. In order to verify the relevance of these factors we apply the segmentation process described in (*Romeo et al. 2012a*), which is based on the computation of a global threshold through a fuzzy clustering learning-based strategy. The images under analysis contain two main classes of interest, i.e., vegetation and no vegetation. The fuzzy clustering is designed to compute the cluster centers associated to each of these two classes based on specific samples of the image. These samples are represented by features with the three RGB spectral components. The cluster centers are three dimensional vectors with three spectral values. Once the cluster centers are estimated, a threshold value is obtained as the percentage of the green spectral component with respect the three ones. This threshold allows us to identify green plants from the remaining parts in the images and the corresponding binary image is finally obtained. *Figures 4.11* displays the results of applying this process to the images displayed in *figures 4.4*, respectively, which were obtained with two different exposure times.



**Figure 4.11:** Segmented images obtained from the ones in figure 4.4 respectively.

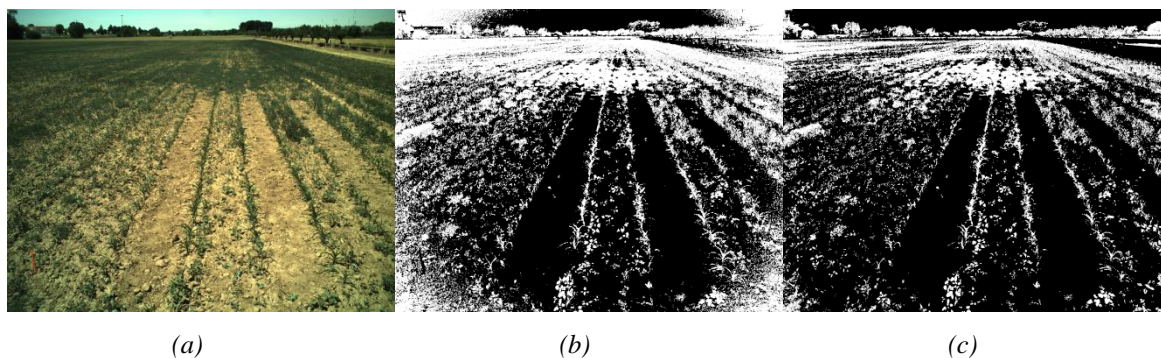
The exposure time for the image in *figure 4.4(a)* was below which it would be desirable, unlike the one obtained in the image in *figure 4.4(b)*, which was sufficient. As we can see, the result in the first case is worse than in the second one. Indeed, the binary image displayed in *figure 4.11(b)* contains an over-segmentation in the part of interest, where crop lines and weeds are placed with important gaps on the outer part. A possible explanation to this phenomenon is that the sensor requires sufficient time to be impacted by the reflectance and the illumination coming from objects in the scene. Because there are different types of materials, the reflectance and illumination sent to the sensors is different for each type of material. When the exposure time is insufficient, the sensor produces this kind of effect. On the contrary, if the exposure time is excessive the intensity image becomes saturated and the image segmentation process fails. *Figure 4.12(a)* displays a saturated image and its corresponding segmented image in *figure 4.12(b)* using the same segmentation procedure as before; we can see how the result becomes unfeasible. From the point of view of weeds and crop lines detection, this leads to clear inaccuracies.

From the point of view of accuracy, color or white balancing is not required because these processes are only suitable for a correct human perception (*Cohen 2013*). This represents an important advantage over other systems that perform this kind of operations.



**Figure 4.12:** (a) Original saturated image; (b) binary image after segmentation.

Regarding the process related to *vignetting*, *figure 4.13(a)* displays an original image and *figure 4.13(b, c)* the binary segmented images without and with *vignetting* correction. In *figure 4.13(b)* we can easily see how an excess of white pixels appears at the four corners representing green plants, which is not present in the image of *figure 4.13(c)* after *vignetting* correction. Thus, when no *vignetting* correction is applied, high inaccuracy results in the corners during weed and crop line detection.



**Figure 4.13:** (a) Original image; (b) binary image without vignetting correction; (c) binary image with vignetting correction.

In order to test the performance of the proposed method to control the adverse illumination factors, we have applied the following procedure:

1. With the tractor stopped at different positions in the field, we acquire an image that needs exposure time adjustment according to the procedure described in *section 4.3.2. Accuracy from the Point of View of Illumination Factors*. A second image is acquired after the adjustment. Thus, for each

position two images are acquired and stored as RGB in the BMP format, *i.e.*, without compression to avoid losses.

2. Each image of the stored pair is processed with and without *vignetting* correction. Thus, for each pair we obtain four binary images through the method described in (Romeo *et al.* 2012a), as already mentioned.
3. For each pair of images a ground-truth image is built as follows. From the four binary images obtained above, we select the one with the highest quality according to an expert human criterion by comparing it with the pair of original images. The selected binary image is manually touched up, so that isolated or groups of pixels are relabeled as white or black pixels, also according to the human criterion.
4. The four binary images are compared against the corresponding ground-truth by computing the Correct Classification Percentage (PCC) index (Rosin and Ioannidis 2003):

$$PCC = \frac{TW + TB}{TW + TB + FW + FB} \quad (4.3)$$

where  $TW$  (true whites) and  $TB$  (true blacks) are the number of white/black pixels respectively in the image that are also white/black in the ground-truth;  $FW$  (false whites) and  $FB$  (false blacks) are the number of white/black pixels respectively in the image that are also black/white in the ground-truth.

We have analyzed 25 pairs of images captured as described in *Step 1* above during different days and under different illumination conditions. *Table 4.2* displays the averaged PCC values over the 25 binary images obtained after the processing with and without exposure time adjustment and *vignetting* correction. The analysis is carried out after removing the top third of each image, because this part is out of our specific interest.

	<b>Images Requiring Exposure Time Adjustment</b>		<b>Images after Exposure Time Adjustment</b>	
<b><i>Vignetting correction</i></b>	No	Yes	No	Yes
<b>PCC</b>	66	74	83	91

**Table 4.2:** Averaged PCC values for the images analyzed with and without exposure time adjustment and with and without *vignetting* correction.

As we can see from results in *table 4.2.*, the highest percentage is obtained after exposure time adjustment and *vignetting* correction. This implies that the best accuracies are obtained after the application of these two processes.

## 4.5. Conclusions

The present study is part of the work developed in the RHEA project (*RHEA 2014*), where the reason for developing this study was to acquire images with the highest possible accuracy for weed and crop line detection. We have studied two main sources of inaccuracies. The first could be caused because of an incorrect arrangement of the extrinsic parameters, once the intrinsic ones have been set beforehand due to the different requirements derived from the agronomic application at hand. Some of these requirements are that ROI to be treated must be placed at a certain distance so that we have enough time (real-time) to process the image and posterior action for treatment. On the same way, a very high pitch angle may make some of the crop lines of interest disappear from the image. Another question to be considered in our case is that the ROI cannot be placed at the very top or at the very bottom of the image because once the tractor is in movement, irregularities on the ground produce oscillations of the tractor and the camera that may make the ROI disappear from the image. Additionally, depending on each project, it is obvious that in other projects extrinsic and intrinsic parameters can be combined to achieve a trade-off among them for maximum accuracy. The second source is derived from the uncontrolled illumination that causes insufficient or excessive CCD sensor activation, producing infra- and over-saturation. Also the illumination causes the known *vignetting* effect, when it crosses an UV-IR coated cut-off filter.

We have also proposed solutions for correcting the adverse illumination effects with the goal of maximum accuracy, concluding that for outdoors image processing, illumination must be constantly calibrated in real time as images are taken. There is no “unique” value for exposure time or iris aperture valid for any atmospheric weather conditions.



## A NEW EXPERT SYSTEM FOR GREENNESS IDENTIFICATION IN AGRICULTURAL IMAGES

### 5.1. Introduction

Camera based devices is an excellent sensor for several applications. One of them is in agriculture where autonomous vehicles equipped with cameras are demanding solutions to distinguish plants (crops and weeds) with the aim of applying treatments over site-specific areas in a larger field (*Davies et al. 1998*). Focusing on maize or barley fields, one of the most important treatments is weeds killing. Different methods and strategies for plant identification have been applied in different works (*Onyango and Marchant, 2003; Tellaeché et al., 2008a,b; Burgos-Artizzu et al., 2009; Guijarro et al., 2011; Montalvo et al., 2012; Guerrero et al., 2012*). *Lopez-Granados 2011*, makes a revision of methods where plant identification is a key step in the process. Most existing strategies address the problem of green identification under the assumption that plants display a clear high degree of greenness. The images come from outdoor environments, which are affected for a high variability of illumination conditions: sunny or cloudy days or both with high rate of changes. In sunny days the position of the sun with respect the camera makes the illumination component impact differently and plants have different shades and shadows in the image. Furthermore, the greater the solar illumination the more intense infrared and ultraviolet radiations are. Also, high components of illumination could cause sensor saturation. These situations are absolutely normal in agricultural environments. Although modern camera devices have the ability to make auto image corrections through auto-image white balance or enhancement, most times this is insufficient, particularly in industrial cameras, where most camera settings must be dynamically adjusted (exposure time, auto-image white balance) or fixed in advance by the user (focal length, iris aperture).

This work proposes a new automatic and robust Expert System for greenness identification which is able to cope with the identification of green plants even though adverse environmental conditions. It consists of two main modules: 1) decision making, based on histogram analysis and 2) greenness identification, where two different strategies are proposed, the first based on classical methods and the second inspired on the Fuzzy Clustering approach. The Expert system design as a whole and the Fuzzy Clustering strategy make the contributions of this work. The performance of the method allows verifying its viability for automatic tasks in agriculture, involving the identification of green plants.

This work has been developed within RHEA project (RHEA 2014) in the seventh framework of the European Union of nanoscience and nanotechnologies as well as in the AGL (2011) project. Initial results were published in *Romeo et al. (2013b)*, *Romeo et al. (2012b)*, *Montalvo et al. (2012b)* y *Guijarro et al. (2013)*.

## 5.2. State of the art

Some years ago *Tian and Slaughter (1998)*, considered the images captured under different conditions such as sunny or cloudy days, affecting illumination variability and assuming that they are typical situations in agricultural images coming from outdoor environments. Later several strategies have been proposed for segmenting crop canopy images, specifically oriented towards green segmentation:

1) *Visible spectral-index based*, including the excess green index *ExG* (*Woebbecke et al. 1995; Ribeiro et al. 2005*), the excess red index *ExR* (*Meyer et al. 1998*), the colour index of vegetation extraction *CIVE* (*Kataoka et al. 2003*), the excess green minus excess red index *ExGR* (*Neto 2004*) and the vegetative index *VEG* described in *Hague et al. (2006)*, which is designed to cope with the variability of natural daylight illumination. *ExG*, *ExGR*, *CIVE* and *VEG* have been applied under a combined form in *Guijarro et al. (2011)*, gaining in performance with respect to their individual application. All these approaches need to fix a threshold for final segmentation, i.e. to discriminate between plants and other parts (soil, sky).

2) *Specific threshold-based approaches*, including dynamic thresholding. Generally, these techniques assume a two-class problem where plants and soil are to be identified. *Reid and Searcy (1987)*, estimate a decision function under the assumption

that the classes follow Gaussian distributions. The Otsu's method (*Otsu 1979*) is also applied considering a bi-class problem (*Ling and Ruzhitsky 1996; Shrestha et al. 2004*). These algorithms are applied to gray images. *Gebhardt et al. (2006)*, apply also thresholding for segmentation, and transform the images from RGB to gray scale intensity. This approach was later improved using local homogeneity and morphological operations in *Gebhardt and Kaihbauch (2007)*.

*Kirk et al. (2009)*, apply a combination of greenness and intensity derived from the red and green spectral bands and compute an automatic threshold for a two-class problem assuming two Gaussian probability density functions associated to soil and vegetation respectively; this procedure requires the previous estimation of an angle to rotate the hypothetical greenness axis. *Meyer and Camargo-Neto (2004)* have applied the automatic Otsu's thresholding method for binarizing *ExG* and the normalized difference index *NDI*, where a comparison is established against the segmentation obtained from *ExGR* determining that in this last case, a value of zero suffices for the threshold, therefore the Otsu's method is not required. *Guijarro et al. (2011)* and *Burgos-Artizzu (2011)*, have applied the statistical mean value of the transformed image obtained with the vegetation indices instead of automatic thresholding such as Otsu. They justify its choice because Otsu's method gives a threshold value higher than the mean and produces infra-segmentation, i.e. some plants are not conveniently identified. Thresholding is not exclusive of applications involving green plants (*Moreno-Matías and Gómez-Gil, 2007*).

3) *Learning-based Meyer et al. (2004)*, have applied unsupervised approaches, including fuzzy clustering, for segmenting regions of interest from *ExR* and *ExG*. *Tian and Slaughter (1998)*, proposed the environmentally adaptive segmentation algorithm (EASA) for detecting plants through a supervised learning process. *Ruiz-Ruiz et al. (2009)* applied the EASA under the HSI (hue-saturation-intensity) color space to deal with the illumination variability. (*Zheng et al. 2009-2010*) use a supervised mean-shift algorithm under the assumption that the segmentation of green vegetation from a background can be treated as a two-class segmentation problem; the class separability is validated through a neural network and the Fisher linear discriminant respectively, the color spaces used were RGB, LUV and HSI. *Guerrero et al. (2012)*, apply Support vector machines for the identification of the two clusters.

### 5.3. Motivational research of the proposed strategy

The above methods are intended for plant identification through their greenness, based on the accentuation of the green color (*Meyer and Camargo-Neto 2008*), but their effectiveness drops when, in the plants analyzed, the green spectral component loses its relevance because of adverse environmental conditions where such component takes similar values and sometimes lower than the red one.

Moreover agricultural images contain not only green plants but other structures, (soil, sky, debris, crop residues or shades), that must be conveniently addressed.

The main direct effect caused by the illumination is reflected in the image histogram, where the contrast is a decisive factor for greenness identification. Classical methods (e.g.: *Extra*, *CIVE* or *VEG*) work properly for well-contrasted images but fail miserably when images are insufficiently contrasted although they are later enhanced. This situation occurs most often in images captured with industrial cameras connected to a computer for image processing. Another undesired situation appears when important parts in the images become highly saturated.

Thus, our system is designed with a first decision making module based on image histogram analysis, which determines if the incoming image contains sufficient quality to apply classical greenness identification methods. Otherwise, the image is to be processed by a new greenness strategy, which is a method specifically developed with such purpose.

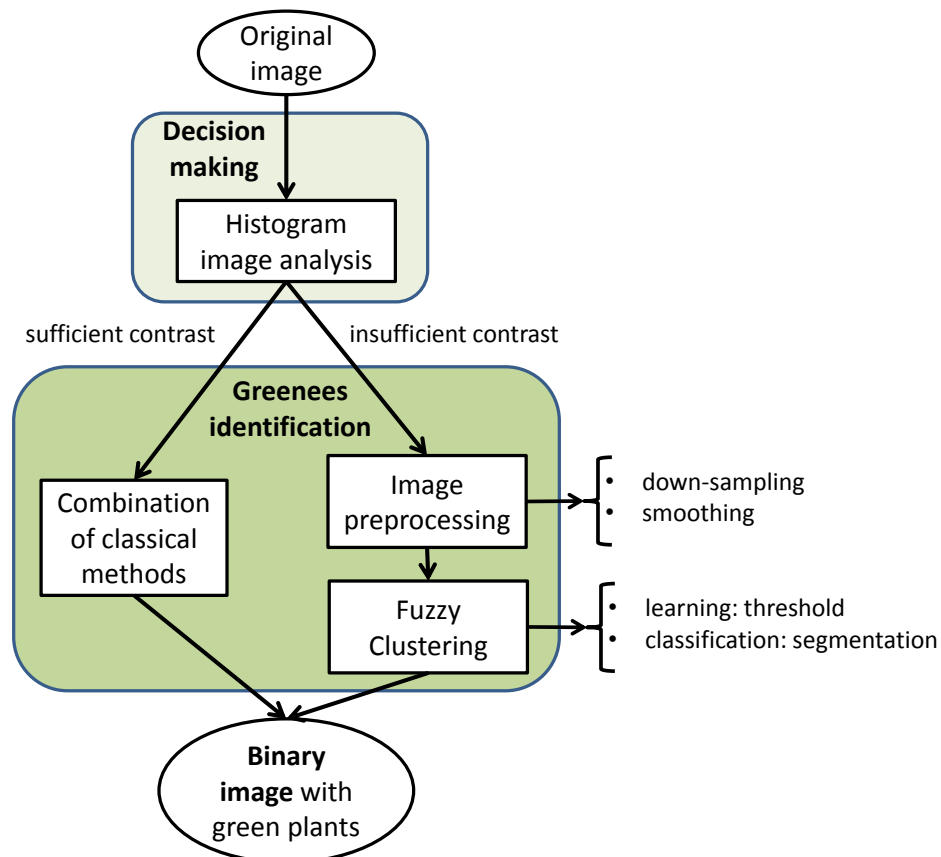
We focus this specific method as a bi-classification approach where green plants are considered as belonging to a class and the remainder elements in the image are assigned to the other class. The Fuzzy Clustering (*FC*) approach is conveniently adapted for this purpose. *FC* consists of two phases, namely: learning and classification. The learning phase is exploited to determine a dynamic threshold for each image and the classification is reduced to a simple decision making process. Previous to this process we apply image down sampling to achieve an image resolution with the aim of saving processing time so that the image fulfills real time requirements if any.

Thus, the idea is to apply an automatic strategy for image segmentation based on the potential ability of the *FC* approach, where its learning phase is exploited to obtain a

specific threshold valid for each image. No learning is required in the general sense; this avoids the need of prior training.

## 5.4. Expert system design

The Expert System (ES) architecture is displayed in *figure 5.1*, where two main modules are identified: *Decision making* and *Greenness identification*. The incoming image is transferred to the decision making module, where based on histogram analysis a decision is made according to the contrast of the histogram. If sufficient contrast, we apply the combined method described in *Guijarro et al. (2011)*, where at the end a binary image is obtained with the green plants identified; otherwise when the image displays what we consider an insufficient contrast, it is firstly image processed by applying a down-sampling together with smoothing. After this preprocessing, a procedure inspired on the *FC* approach allows to determine a threshold, during a phase equivalent to what is considered as learning in the classical *FC* approach. Once the threshold is available, through a simple decision rule, a binary image is obtained, also with the green plants identified. In what follows we explain all process involved in the ES.



*Figure 5.1: Expert System architecture.*

### 5.4.1. Decision making

This module is designed with the aim of analyze the quality of the incoming image based on its histogram. After the analysis, a decision is made and the image is processed according to one of two available processes in the greenness identification module. The histograms for the three spectral channels from the RGB image are processed. Each histogram is defined by its corresponding discrete function  $h(a) = n_a$ , where the parameter  $a$  represents the gray level and  $n_a$  is the number of pixels with gray level  $a$ . In our 8-bit RGB channel representation  $a$  ranges in  $[0, 255]$ . Dividing  $h(a)$  by the total number of pixels in the image, the estimation of the probability is obtained, i.e.  $p(a) = h(a)/M \times N$ , where  $M$  and  $N$  are both sizes of the image. The histogram provides useful image statistics that are exploited for determining the image quality. Let  $a$  be a random variable denoting gray levels, the  $n^{\text{th}}$  moment of  $a$  about the mean is defined as (Gonzalez and Woods 2008):

$$\mu_n(a) = \sum_a (a - m)^n p(a) \quad (5.1)$$

where  $m$  is the mean value of  $a$ , i.e. the average gray level

$$m = \sum_a a p(a) \quad (5.2)$$

The most important statistical parameters from the point of view of image quality are (Holub and Ferreira 2006): mean,  $m$ , variance  $\nu = \mu_2(a)$ , skewness  $\gamma = \mu_3(a) / \mu_3^{3/2}(a)$  and kurtosis  $\kappa = \mu_4(a) / \mu_2^2(a)$ . The mean determines the average level of brightness, where low, high and medium values indicate the degree of light which has impacted the device. The moments provide information about the distribution of values around the mean. Variance is a measure of gray-level contrast, where high values indicate dispersion of values around the mean and low values are indicative of a high concentration of values around the mean. The skewness measures the asymmetry in the distribution. A right skewness is presented when the histogram displays a large tail oriented towards high brightness values and high concentration in the part of low brightness values (positive skewness). In the opposite case the skewness is negative. The kurtosis provides information about the peak in the distribution; low kurtosis indicates flat top parts in the histogram around the mean but high values are indicative

of peaks around the mean with high slopes and large tails. Skewness and kurtosis are both zero for Gaussian distributions.

Once we have defined the above parameters, the question is: how can we determine the image quality and what means quality from the point of view of our agricultural images? The answer comes from the experience in this area. Images highly contrasted are considered as images with sufficient quality and vice-versa. An image with sufficient contrast should be identified by mean values in the central part of histogram, high variance, low skewness (positive or negative) and high kurtosis. On the contrary, an image with insufficient contrast is identified by mean values either low or high, high skewness (positive or negative) and low kurtosis. The next step is to determine the ranges of variability for the above parameters and accordingly the derivation of the corresponding decision rule. With such purpose we have analyzed a set of images described in *section 5.5. Results*, such images are classified as images with sufficient and insufficient contrast from which we obtain their mean, variance, skewness and kurtosis values. A set of three rules, each with four premises suffices for making decisions

$$\begin{aligned} \text{If } m_{ai} < m_i < m_{bi} \text{ y } v_i > v_{ai} \text{ y } |\gamma_i| < |\gamma_{ai}| \text{ y } \kappa_i > \kappa_{ai} &\rightarrow \text{sufficient contrast} \\ \text{otherwise} &\rightarrow \text{insufficient contrast} \end{aligned} \quad (5.3)$$

where  $i = R, G, B$ ; i.e. three values of each of the three parameters are obtained according to the three RGB spectral channels. Parameters with subindices  $a$  and  $b$  are upper and lower limits to be fixed in *section 5.5*.

### 5.4.2. Greenness identification

According to the scheme displayed in *figure 5.1.*, if the incoming image contains sufficient contrast a combination of well-tested methods is applied based on the experiments carried out in *Guijarro et al. (2011)*. Otherwise, a new strategy inspired on the *FC* approach is proposed for image thresholding.

### 5.4.2.1. Combination of methods

Given an original input image in the RGB color space, we apply the following normalization scheme, which is usually applied in agronomic image segmentation (*Gée et al. 2008*):

$$r = \frac{R_n}{R_n + G_n + B_n}$$

$$g = \frac{G_n}{R_n + G_n + B_n} \quad (5.4)$$

$$b = \frac{B_n}{R_n + G_n + B_n}$$

where  $R$ ,  $G$  and  $B$  are the normalized RGB coordinates ranging from 0 to 1 and are obtained as follows:

$$x = \frac{x}{x_{max}} \quad (5.5)$$

where  $x = R, G, B$  and  $x_{max} = 255$  for our 24-bit color images. Vegetation indices to be combined are computed as follows (see references above in *section 5.2*).

$$\text{Excess green, } ExG = 2g - r - b \quad (5.6)$$

$$\text{Excess green minus excess red, } ExGR = ExG - 1.4r - g \quad (5.7)$$

$$\text{Color index of vegetation extraction, } CIVE = 0.441r - 0.811g + 0.385b + 18,78745 \quad (5.8)$$

$$\text{Vegetative, } VEG = \frac{g}{r^a b^{1-a}} \quad (\text{con } a = 0,667 \text{ según } Hague \text{ y col. } 2006) \quad (5.9)$$

Based on *Guijarro et al. (2011)*, the above three indices are combined to obtain the resulting value  $COM$  as follows,

$$COM = w_{ExG} ExG + w_{ExGR} ExGR + w_{CIVE} CIVE + w_{VEG} VEG \quad (5.10)$$

where,  $w_{ExG}$ ,  $w_{ExGR}$ ,  $w_{CIVE}$  and  $w_{VEG}$  are weights for each index, representing the relative importance of the index. *Guijarro et al. (2011)* provide the values for the four weights

participating in the combination, which are the following:  $w_{ExG} = 0,25$ ,  $w_{ExGR} = 0,30$ ,  $w_{CIVE} = 0,33$  and  $w_{VEG} = 0,12$ .

The resulting combined image  $COM$ , is linearly mapped to range in  $[0, 1]$ , after which, it is thresholded by applying the Otsu's method, obtaining a binary image, where white pixels identify plants in the original image, with clear spectral RGB components associated to unmasked plants. On the contrary, black pixels identify those pixels in the original image belonging to masked plants, soil and other materials present in the field.

#### **5.4.2.2. Preprocessing and Fuzzy Clustering**

Two consecutive strategies are proposed when the images are insufficiently contrasted, i.e. image preprocessing and  $FC$ . The former is proposed mainly for image reduction, although it could be avoided if restrictive real-time problems exist. Indeed, because  $FC$  involves a first phase associated to a learning process (iterative) the number of samples, which are pixels coming from the image, determine the computational cost. The lower the number of pixels to process lower the computation time. The second ( $FC$ ) is the kernel of this process, where after the learning phase a threshold is obtained which allows the image binarization.

##### *a) Image preprocessing*

The original image is down sampled by applying successive Discrete Wavelets Transform ( $DWT$ ) (*Pajares and Cruz 2004*). Only the  $DWT$  is of interest applied over the approximation coefficients at each resolution level and for each spectral R, G, B channel. In this way, as the same time the signal is decomposed, a smoothing process is applied to minimize possible noise or spurious pixels caused by the adverse environmental outdoor conditions.

##### *b) Image thresholding by Fuzzy Clustering*

*b.1) Learning phase and thresholding:* given the down sampled RGB image, it contains  $n$  pixels which are stored in  $X$ , i.e.,  $X = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathcal{R}^d$ , where  $d$  is the data dimensionality. Therefore, each sample vector  $\mathbf{x}_i$  represents an image pixel, where its components are the three RGB spectral components of that pixel at the original image location  $(x, y)$ . This means that in our experiments the data dimensionality is  $d = 3$ . Each sample is to be assigned to a given cluster  $w_j$ , where the number of possible

clusters is  $c$ , i.e.:  $j = 1, 2, \dots, c$ . In the proposed approach  $c$  is set to 2 because we were only interested on two types of textures, i.e. green plants (crop/weeds) and the remainder (soil, debris, stones).

The samples in  $X$  are to be classified based on the well-known fuzzy clustering approach that receives the input training samples  $\mathbf{x}_i$  and establishes a partition, assuming the number of clusters  $c$  is known. The fuzzy partitioning problem for  $b > 1$  is minimize the following variance criterion,

$$\min J_b(\mu_{ij}, \mathbf{v}, t) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij}(t))^b \|\mathbf{x}_i - \mathbf{v}_j\|^2 \quad (5.11)$$

The process computes for each  $\mathbf{x}_i$  at the iteration  $t$ , its degree of membership in the cluster  $w_j$  ( $\mu_i^j$ ) and updates the cluster centers  $\mathbf{v}_j$  as follows (*Zimmermann H.J. 1991*):

$$\mu_i^j(t+1) = \frac{1}{\sum_{r=1}^c \left( \frac{d_{ij}(t)}{d_{ir}(t)} \right)^{\frac{2}{b-1}}} \quad (5.12)$$

$$\mathbf{v}_j(t+1) = \frac{\sum_{i=1}^n [\mu_i^j(t)]^b \mathbf{x}_i}{\sum_{i=1}^n [\mu_i^j(t)]^b} \quad (5.13)$$

$d_{ij}^2 \equiv d^2(\mathbf{x}_i, \mathbf{v}_j)$  is the squared Euclidean distance. The number  $b$  is called the exponential weight (*Bezdek 1981 and Duda et al. 2000*),  $b > 1$ . The stopping criterion of the iteration process is achieved when  $\|\mu_i^j(t+1) - \mu_i^j(t)\| < \varepsilon \forall ij$  or a number  $t_{max}$  of iterations is reached.

The method requires the initialization of the cluster centers, so that the equation (5.12) can be applied at the iteration  $t = 1$ . Because our interest consists in the identification of green plants among other parts, one ideal center should be the one with the pure green color, without loss of generality this center is initially set with the following values  $\mathbf{v}_1 = (0, 255, 0)$ ; the second center only requires that the green spectral

component is not dominant over the red and blue, the following values suffice for it,  $\mathbf{v}_2 = (255, 0, 128)$ .

Once the learning process is finished we obtain two cluster centers  $\mathbf{v}_1$  and  $\mathbf{v}_2$  associated to clusters  $w_1$  and  $w_2$  with components  $\mathbf{v}_1 = (v_{1R}, v_{1G}, v_{1B})$  and  $\mathbf{v}_2 = (v_{2R}, v_{2G}, v_{2B})$  respectively; where  $v_{1R}$ ,  $v_{1G}$  and  $v_{1B}$  or  $v_{2R}$ ,  $v_{2G}$  and  $v_{2B}$  represent the averaged values for the corresponding RGB spectral components at each cluster obtained according to equation (5.12). Thus, because  $\mathbf{v}_1$  is associated to the cluster where green plants belongs to, we compute the threshold value  $T = v_{1G}/(v_{1R}, v_{1G}, v_{1B})$ , which represents the proportion of greenness with respect the three spectral components. It is expected green plants tend to have their green spectral components in the RGB color model the greatest. Thus we chose  $T$  as the threshold for the posterior decision making process.

*b.2) Classification phase and segmentation:* Given the original RGB image, we obtain the three  $r$ ,  $g$  and  $b$  normalized values through equation (5.5). Considering the threshold  $T$  computed as above, the following rule allows obtaining a binary image,  $W$ , where white pixels are identified as green pixels in the original image:

$$W = \begin{cases} 1 & \text{si } \frac{g}{r+g+b} > T_G \\ 0 & \text{en caso contrario} \end{cases} \quad (5.14)$$

## 5.5. Results

The images used for this study were acquired in La Poveda Research Station, Arganda del Rey, Madrid at the Spanish Research Council facilities and also in simulated fields at the region of the Alcarria, Trijueque (Guadalajara). They were taken with the following digital color camera devices and dates: a) HPR817 with image resolution of  $600 \times 800$  in March/April/May 2007 (330 images) for barley and maize crops; b) Canon EOS 400D with resolution of  $1944 \times 2592$  during April/May 2011 (350 images); c) Basler 17FC 1400 connected to a laptop through FireWire IEEE 1394 and resolution of  $1392 \times 1038$  in April/May 2011 (320 images); d) Kodak EasyShare M380 with resolution  $1280 \times 960$  in March 2011 for barley crop and May 2012 for maize crop

(210 images); e) SVS-VISTEK with GigaEthernet connection to a laptop and resolution of  $2336 \times 1752$  equipped with UV-IR cut Schneider filter 486 in March/April/May 2012 (340 images) for barley and maize crops. This makes a total of 1550 images.

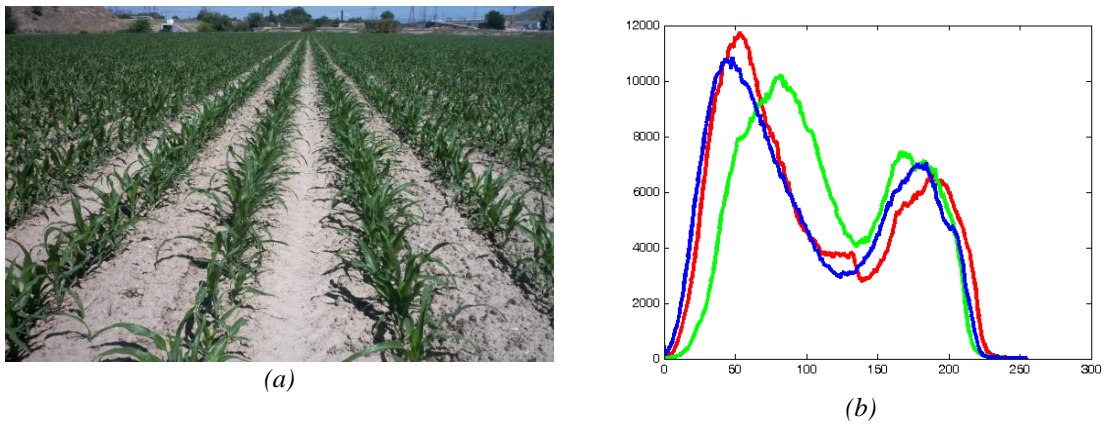
All acquisitions for the same crop were spaced by different days varying from five/six/ten until thirty days, under a high variety of illumination conditions, including: cloudy and sunny days with sun incident around all angles with respect the camera; during the morning, afternoon, midday and late afternoon.

These digital images were captured under perspective projection in barley and maize fields containing soil with different elements (stones, debris or old crop residues), plants (crop and weeds), sky (with and without clouds) and other elements such as buildings. They were stored as 24-bit color images and saved in RGB (Red, Green and Blue) color space in the BMP and JPG formats. The expert system was implemented in Matlab R2009b (*The Mathworks 2014*) and the images were processed with its Image Processing Toolbox.

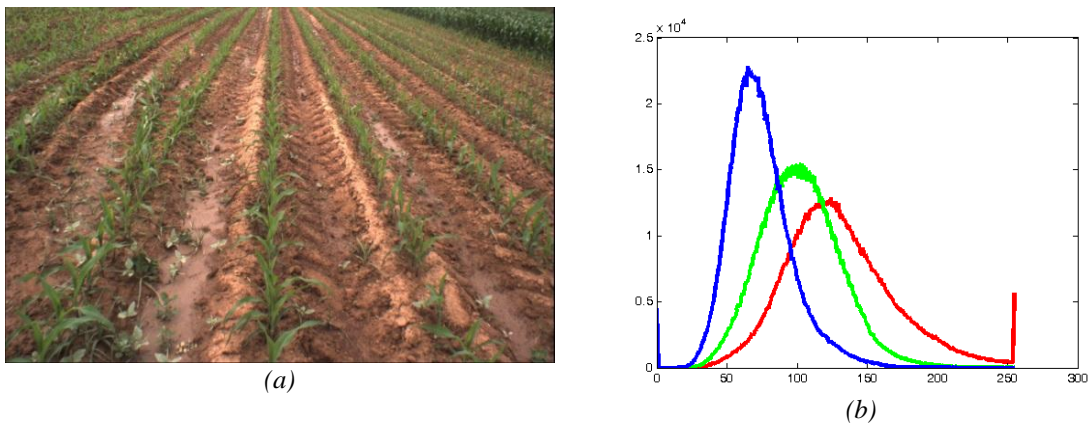
### 5.5.1. Decision making module: identification of parameters

From the set of available images, we have randomly selected the 5% of images obtained with each camera device, including images of barley and maize when available.

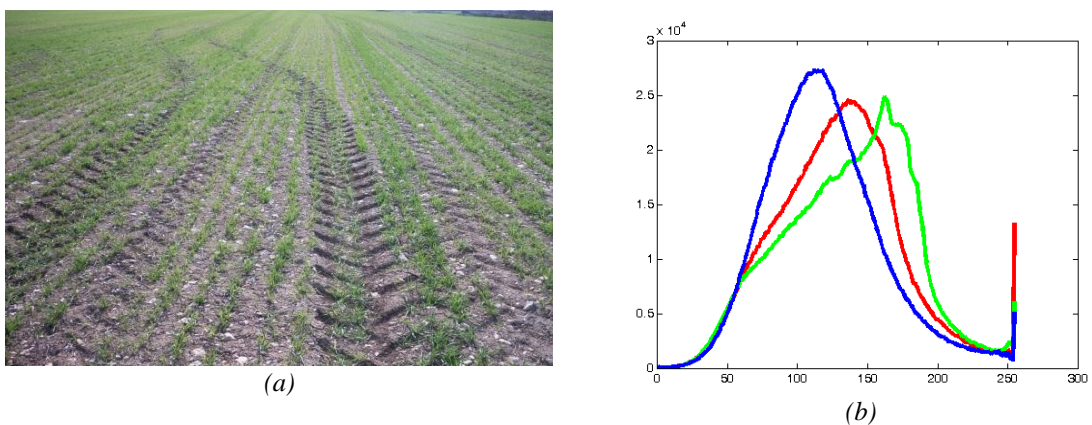
According to the Greenness identification module in *figure 5.1*, for each image we obtain two binary images, one by applying the combination of classical methods according to equation (5.10) and the second by applying image preprocessing and Fuzzy Clustering, equations (5.11) and (5.12). The binary images are visually analyzed by an expert in order to determine the best result, considering how the green plants, present in the original images, have been identified. *Figures 5.2(a) to 5.4(a)* display three representative images where the best binarization is achieved by applying combination of classical methods. On the contrary, *figures 5.5(a) to 5.7(a)* display three representative images in which the best binarized results have been obtained with the processing involving *FC*, i.e. they are considered images with insufficient contrast. *Figures 5.2 to 5.7* labeled with *(b)* display the corresponding histograms for the three RGB spectral channels.



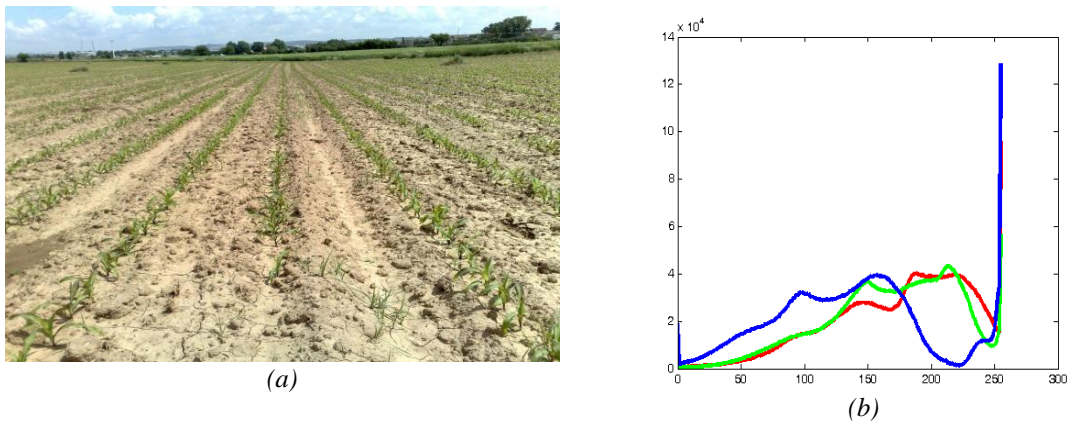
**Figure 5.2:** (a) Original image (maize) captured with the Kodak EasyShare M380 camera; (b) Histogram for the three RGB spectral channels.



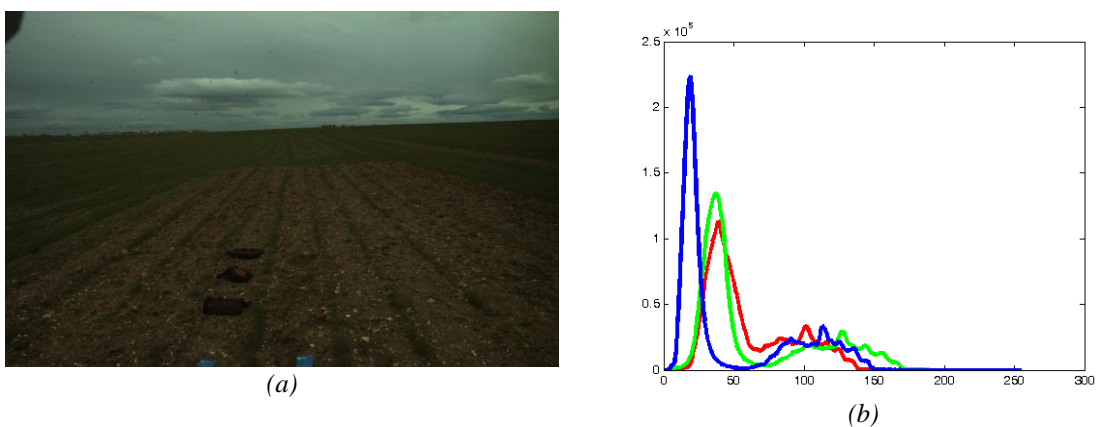
**Figure 5.3:** (a) Original image (maize) captured with the 17FC 1400 Basler camera; (b) Histogram for the three RGB spectral channels.



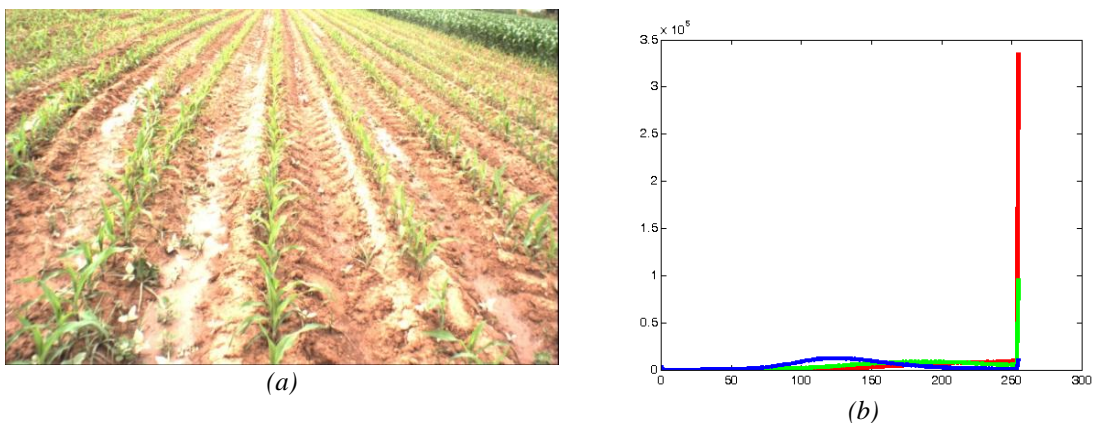
**Figure 5.4:** (a) Original image (barley) captured with the HPR817 camera; (b) Histogram for the three RGB spectral channels.



**Figure 5.5:** (a) Original image (maize) captured with Canon EOS 400D; (b) Histogram for the three RGB spectral channels.



**Figure 5.6:** (a) Original image (barley with wide row crops simulating maize) captured with SVS-VISTEK camera; (b) Histogram for the three RGB spectral channels.



**Figure 5.7:** (a) Original image (maize) with the 17FC 1400 Basler; (b) Histogram for the three RGB spectral channels.

Figures 5.2 to 5.4 display images with histograms with distributions covering the whole range of values and with maximum values located at the central part. On the contrary, histograms associated to figures 5.4 to 5.7 are heeled left or right. This represents an important finding, which is decisive for making a decision. The only question is to quantify this observation in some way.

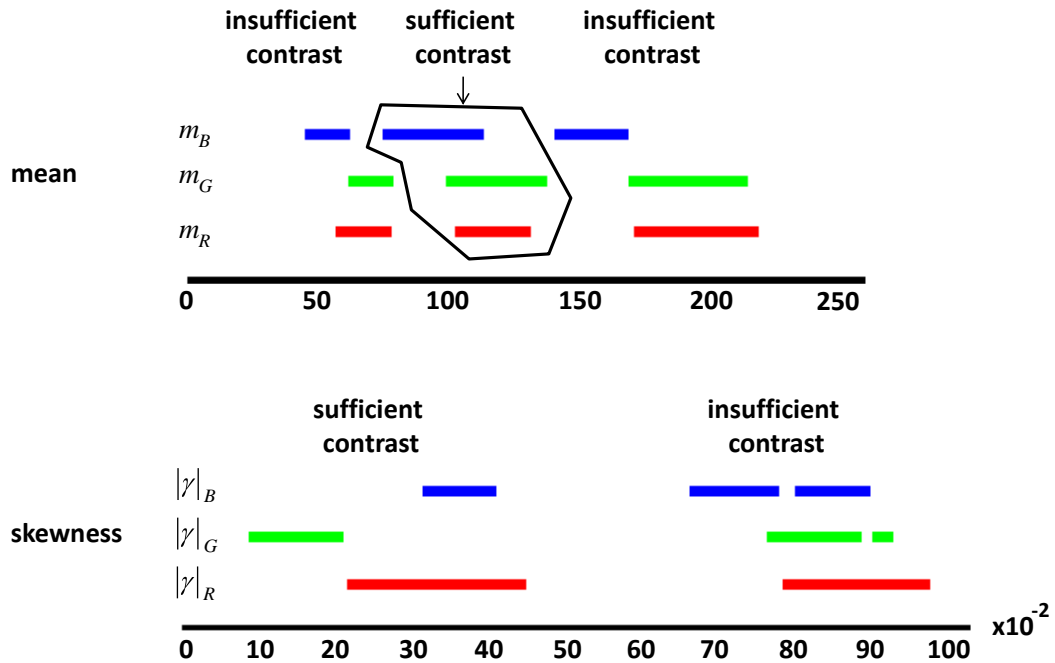
Table 5.1 displays the average parameter values derived from the corresponding histograms on the set of images analyzed according to the best binary image obtained. Minimum and maximum values for each R, G, and B spectral channel are identified for the mean ( $m$ ), variance ( $\nu$ ), skewness ( $|\gamma|$ ) and kurtosis ( $\kappa$ ). They are distinguished as images with sufficient contrast and images with insufficient contrast and inside these last ones we still can distinguish between images with the histogram heeled left or right, this is because the mean histogram values ranges differently.

	Sufficient contrast		Insufficient contrast			
			Left heeled		Right heeled	
	Minimum (R,G,B)	Maximum (R,G,B)	Minimum (R,G,B)	Maximum (R,G,B)	Minimum (R,G,B)	Maximum (R,G,B)
$m$	(106, 102, 78)	(130, 136, 112)	(60, 65, 49)	(75, 81, 60)	(174, 168, 144)	(217, 213, 167)
$\nu \times 10^3$	(1.51, 0.83, 0.52)	(3.73, 2.64, 3.65)	(0.91, 1.83, 1.85)	(2.91, 3.55, 3.89)	(1.83, 2.20, 1.72)	(2.72, 2.53, 3.33)
$ \gamma $	(0.22, 0.10, 0.33)	(0.45, 0.21, 0.41)	(0.85, 0.89, 0.82)	(0.89, 0.93, 0.90)	(0.79, 0.78, 0.68)	(0.98, 0.87, 0.78)
$\kappa$	(1.75, 1.84, 1.72)	(3.69, 3.73, 4.50)	(2.46, 2.26, 2.10)	(4.45, 3.98, 4.60)	(2.77, 2.74, 2.69)	(5.42, 3.08, 3.66)

**Table 1:** Histogram parameters (mean, variance, skewness, kurtosis) identifying minimum and maximum values for images with sufficient/insufficient contrast in the three R, G and B spectral channels.

From results in table 5.1., two important conclusions can be extracted. There are two parameters which are really discriminating ( $m$  and  $|\gamma|$ ) and two that are not ( $\nu$  and  $\kappa$ ). Indeed, considering the Red spectral channel, we can see that the mean for images with sufficient contrast ranges in [106, 130] and for images with insufficient contrast there are two intervals, one for left heeled [60, 75] and one for right heeled [174, 217]. These intervals clearly appear as no overlapped with wide margins between them. This reasoning can be extended to the Green and Blue spectral channels, where this same behavior appears, i.e. [102, 136] and [65, 81]-[168, 213] for Green and [78, 112] and [49, 60]-[144, 167] for Blue. With respect skewness and the Red spectral channel we can see that images with sufficient contrast the range is [0.22, 0.45] and with insufficient contrast [0.85, 0.89] and [0.79, 0.98], i.e. with overlapping between these intervals. Similar behavior is applicable for the Green spectral channel with [0.10, 0.21] and [0.89, 0.93]-[0.78, 0.87] and for the Blue channel is [0.33, 0.41] and [0.82, 0.90]-

[0.68, 0.78]. *Figure 5.8* clarifies this situation, where we can verify indeed that not overlapping exists on each for the averaged mean and skewness values. Each channel is identified with the subindices R, G and B accordingly.



*Figure 5.8:* Representation of averaged values for mean and skewness for the three RGB spectral channels.

The behavior displayed by mean and skewness does not occur for variance and kurtosis where overlapping occurs and therefore, these last ones are not discriminant. Hence only setting of mean and skewness parameters involved in equation (5.3) for each spectral channel (R,G,B) is possible as follows:  $m_{aR} = 106$ ;  $m_{aG} = 102$ ;  $m_{aB} = 78$ ;  $m_{bR} = 130$ ;  $m_{bG} = 136$ ;  $m_{bB} = 112$ , we can assume a certain margin of tolerance because there are sufficient distances for no overlapping; for skewness it is clear that appropriate values for the three spectral channels should be between 0.50 and 0.60, i.e.  $\gamma_{aR} = \gamma_{aG} = \gamma_{aB} = 0.55$ . With respect variance and kurtosis they are no longer considered and they can be removed from equation (5.3).

### 5.5.2. Greenness Identification Module

Once the above parameters have been established, we have randomly selected the 15% of available images, i.e. 232, about half with enough contrast and the other half

with insufficient contrast, which are all different from the ones used for parameter's identification. The performance of this combined strategy and also the one based on *FC* is tested against the corresponding ground-truth images which are built following the procedure described in *Guijarro et al. (2011)*, which is summarized here for facility as follows: (1) given the original image in the RGB color space, we extracted all pixels where the Green component is greater than the Red and Blue ones simultaneously at least in a 10%. This allows us to obtain an important part of green plants, which are labeled as white; (2) also from the original image we extracted those parts where the Red and Blue components are dominant, i.e. they overpass at least in a value of 10% the other two. This allows extracting the Red and Blue parts with a certain degree of confidence, being labeled as black; (3) the pixels that still were not labeled were manually touched up and labeled according to the human expert criterion by comparing them with the original image. Nevertheless, when the images are highly complex, such as the one in *figure 5.6(a)* or extremely complex as the one displayed in *figure 5.10(a)*, we select parts containing green plants and other elements which are easily distinguishable.

*a) Combination of classical methods*

We have applied exactly the parameters proposed in *Guijarro et al. (2011)* obtaining similar results.

*b) Image preprocessing*

We have applied successive *DWT* decomposition. As mentioned before, this procedure involves both down-sampling and filtering processes. Images with maximum resolution used in this work are those provided by Canon EOS 400D camera, i.e.  $1944 \times 2592$ . We have verified that reductions of  $\frac{1}{16}$  suffices for these images to achieve appropriate performances in terms of effectiveness, as compared to the corresponding ground-truth, while reducing the computational cost during this process and also during the subsequent *FC* process because the number of samples to be processed. This leads to resolutions of  $121 \times 162$  i.e. the number of samples for our experiments is finally fixed to 19602. For original images with resolutions other than the above, the *DWT* decomposition level should be the appropriate to achieve a similar number of samples. Regarding the family of wavelets used, we have verified that this aspect is not relevant,

thus we have used the one that involves less computational cost. The Haar transform fulfill this requirement because its associated low pass filter contains the less number of coefficients and the number of computational operations required for filtering is less.

*c) Fuzzy clustering*

As mentioned above, this process starts with the cluster centers already initialized, then applying successive iterations we estimate the final cluster centers, equation (5.12), until convergence through  $\varepsilon$  or a maximum number of iterations is reached  $t_{max}$ .

In order to fix these parameters we have selected 10 images identified as images with insufficient contrast and its corresponding ground-truth. Initially we fix  $t_{max} = 100$ , which is a number relatively high. We vary  $\varepsilon$  from  $10^{-2}$  to  $10^{-5}$  in steps of  $5 \times 10^{-5}$ , once we achieve the best results according to the ground-truth for each image, we annotate the number of iterations and the  $\varepsilon$  value for each image. *Table 5.2.* displays these values for the ten images tested.

# image	1	2	3	4	5	6	7	8	9	10
iterations ( $t$ )	3	8	12	2	3	1	6	7	3	2
$\varepsilon \times 10^{-5}$	6	11	14	77	73	24	18	10	12	15

**Table 5.2:** Iterations and  $\varepsilon$  values for ten images analyzed

With values in *table 5.2.*, we finally fix  $t_{max} = 14$  and  $\varepsilon = 10^{-5}$ , which guarantee a certain degree of confidence, because they are the most restrictive. These values are the ones used for posterior analysis.

*d) Global performance: quantitative analysis*

The proposed ES is quantitatively analyzed with 222 images with their corresponding ground-truth, where the ones used above for fixing  $t_{max}$  and  $\varepsilon$  are now discarded.

Because these images have been previously identified as images with sufficient (108) and insufficient contrast (114) by the expert, the performance of the *decision making* module is evaluated based on the following magnitudes:

- SC: number of images identified correctly with sufficient contrast

- IC: number of images identified correctly with insufficient contrast
- SCE: number of images identified with sufficient contrast being of insufficient contrast (error)
- ICE: number of images identified with insufficient contrast being of sufficient contrast (error)

From these quantities we compute the percentage of correct classification as,

$$PCC = \frac{SC + IC}{SC + IC + SCE + ICE} \quad (5.15)$$

Finally, according to the decision rule in equation (5.3) excluding variance and kurtosis as explained, we obtained  $SC = 101$ ,  $IC = 109$ ,  $SCE = 7$ ,  $ICE = 5$ , i.e.  $PCC = 0.95$  which is an acceptable result.

The combination of classical methods is tested with the SC and ICE images, from which we have achieved the 91% as percentage of success, averaged over the number of images tested and also based on the corresponding ground truth, which is a similar result to the one reported in *Guijarro et al. (2011)*.

The performance of our proposed *FC* method is tested with the *IC* and *SCE* images obtaining a percentage of 89.9% of success.

*Table 5.3* displays the averaged CPU computational times for each process involved in the ES for the set of images processed. The value in *FC* is averaged over the number of iterations, i.e. it represents the average time per iteration.

	<b>Histogram analysis</b>	<b>COM</b>	<b>DWT</b>	<b>FC (per iteration)</b>
<b>CPU time (s)</b>	0.19	0.72	1.01	0.03

*Table 5.3: Averaged CPU computational times in seconds for the process involved in the ES.*

Based on results in *table 5.3.*, when an image is identified with sufficient contrast the total CPU time is 0.91s; when *the* image is identified as with insufficient contrast and in the worst case,  $t_{max} = 14$ , the total time for *FC* is 0.6s. This means that

the total time for the full process in this case is 1.62s. Common agricultural vehicles working in agricultural tasks and requiring green identification, navigate at speeds between 4-6 Km/h (1.1-1.7 m/s). Thus, assuming that vision systems, on board these vehicles, have the ability to cover areas of 4 meters long without any difficult we have a window between 3.60 to 2.35s to process the image, which is sufficient according to times above.

e) *Global performance: qualitative analysis*

Figures 5.9(a) to 5.11(a) display representative original images from the set of available images of maize fields; in (b) appears their histograms. Results and conclusions extracted from this analysis are valid for the full set of images tested. In Figure 5.10(a) we can see a limited field area with different weeds densities in the inter-crop rows spaces; alternatively appear low/high densities, which were manually prepared by farmers for testing purposes.

Table 5.4. displays the relevant parameters (mean,  $m_i$ , skewness,  $\gamma_i$ , cluster center  $\mathbf{v}_1 \equiv v_{1R}, v_{1G}, v_{1B}$  and threshold  $T$ ) to make a decision about the process to apply, i.e. *COM* or *FC*; for each image we identify the camera device with which it has been captured.

	<b>Figure 5.9.</b> (Kodak)	<b>Figure 5.10.</b> (SVS-VISTEK)	<b>Figure 5.11.</b> (Basler)
$(m_R, m_G, m_B)$	(116, 124, 105)	(52, 53, 33)	(236, 230, 187)
$(\gamma_R, \gamma_G, \gamma_B)$	(0.32, 0.17, 0.38)	(0.68, 0.89, 1.65)	(-2.74, -2.13, -1.43)
$\mathbf{v}_1 \equiv v_{1R}, v_{1G}, v_{1B}$	(71.2, 91.3, 57.1)	(32.0, 34.6, 19.9)	(176.8, 155.0, 107.2)
$T$	0.42	0.40	0.35

**Table 5.4:** Histogram parameter values (mean, skewness, cluster center  $\mathbf{v}_1$  and threshold  $T$ ) for the images displayed in figures 5.9., 5.10 .and 5.11.

From results in *table 5.4* we can see that according to the decision rule in equation (5.3), image in *figure 5.9*. is identified as belonging to the set of images with sufficient contrast. It should be processed by applying the *COM* procedure according to equation (5.10). Images in *Figures 5.10* and *5.11* are identified as belonging to the set of images with insufficient contrast; they should be processed with the *FC*-based procedure. According to the expert criterion the three decisions made for these images are correct and the histograms distributions confirm these decisions. An important aspect to remark is the variability of the threshold  $T$  which in images with insufficient contrast varies from 0.34 to 0.52 depending on the type of image. Thus, fixing a value as a universal threshold becomes unfeasible. This makes an important contribution of this work, because we provide a procedure for adjusting a dynamic threshold adapted to the image quality under processing, measured by its contrast. Additionally, some relevant information can be extracted from the histogram. This is the case for the image in *figure 5.11(a)* and its histogram in *(b)*, where one can easily infer that this image presents oversaturation. Although the proposed *FC* based approach provides acceptable results which allow to work under critical adverse conditions, perhaps the best alternative should be to control the input of illumination either applying iris modification or decreasing the exposure time in the sensor device. This makes an additional contribution.

Only for comparison purpose we have processed the three images with both *COM* and *FC*, obtaining respectively the corresponding binary images displayed in *Figures 5.9* to *5.11* labeled with *(c)* and *(d)*. Also, for the same purpose, we display in *figures 5.9* to *5.11* labeled with *(e)* the clustering of the samples with the cluster centers identifying the clusters with green plants and the remainder with other elements (soil, sky).

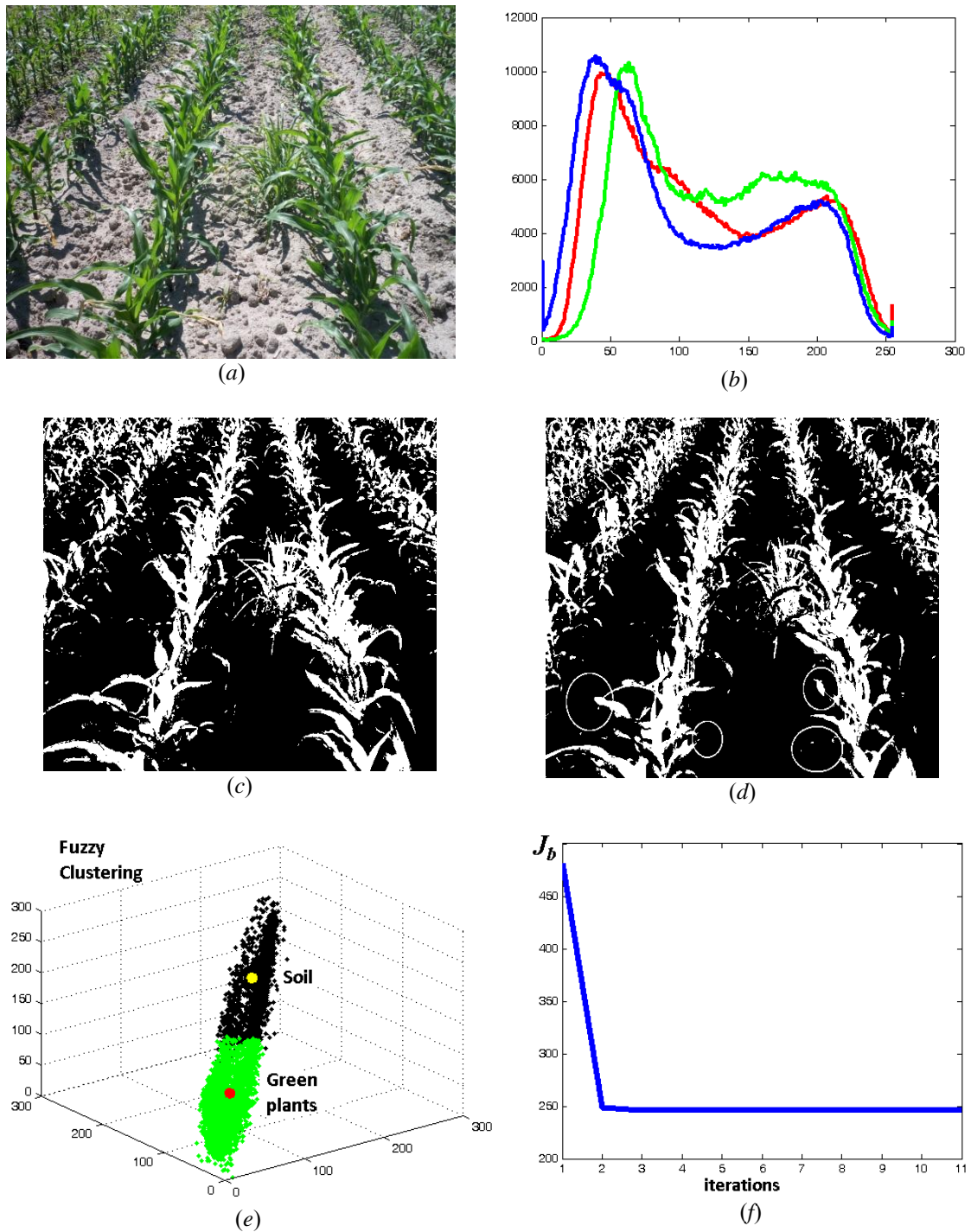
Finally, in figures with labels *(f)* it is displayed the variation of the criterion function defined in equation (5.10) against the number of iterations. It is important to remark once again that during the ES normal operation images with sufficient contrast are not processed under the *FC* approach and with insufficient contrast are not processed with *COM*.

- Analysis of binary images: by comparing the binary images obtained with *COM* and *FC*, we can easily verify the outperformance of the appropriate procedure

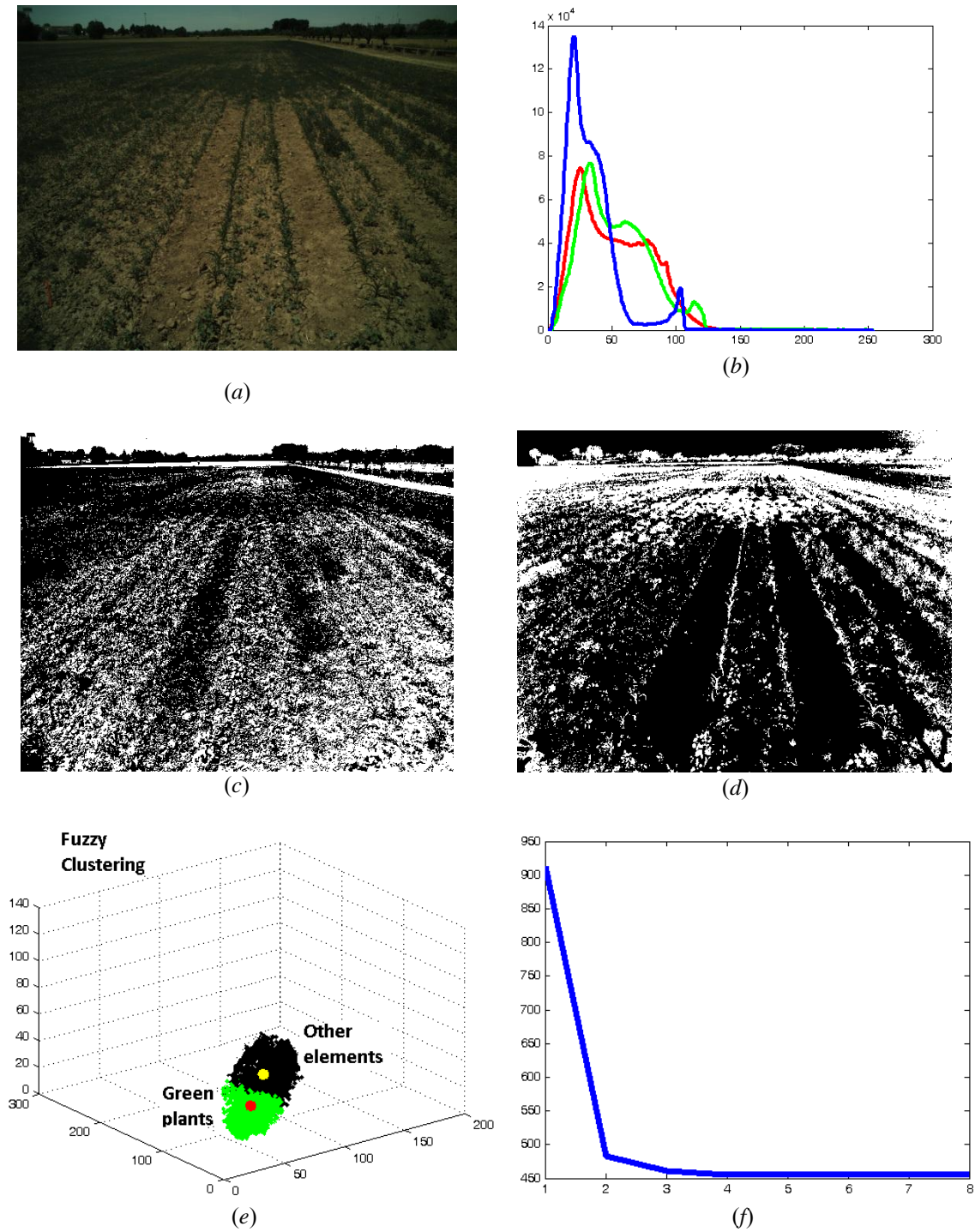
according to the decision made. Indeed, the quality of the binary image in *figure 5.9(c)* is better than the one in *figure 5.9(d)* where important deficiencies are identified with circles, i.e. real plants which have not been identified. This means that *COM* outperforms *FC* as expected. On the contrary, the quality of binary images in *figures 5.10. (d)* and *5.11(d)* is clearly superior to the one obtained in binary images in *figures 5.10. (c)* and *5.11(c)*. Now, this implies that *FC* for images with insufficient quality is the appropriate procedure but not *COM*. These results confirm in a qualitative way that the proposed ES is appropriate for identifying greenness in agricultural images, captured even under adverse environmental conditions.

- Analysis of *FC* (clustering and convergence): *figures 5.9(e)* to *5.11(e)* display the clustering of green plants and the remainder elements in the image, basically soil, sky, trees and some building. As we can see in the three figures, samples belonging to each cluster appear grouped together. Perhaps a greater compactness can be seen in images with insufficient contrast, *figures 5.10(e)* and *5.11(e)*, which is a direct consequence of the histograms and their concentration either left or right. A different measure allows us to verify that in images with insufficient contrast, *figures 5.10(e)* and *5.11(e)*, the Euclidean distance between cluster centers is less than in images with sufficient contrast, *figure 5.9(e)*, this could serve as an additional measurement to verify that the image has been correctly processed if this distance is sufficiently small.

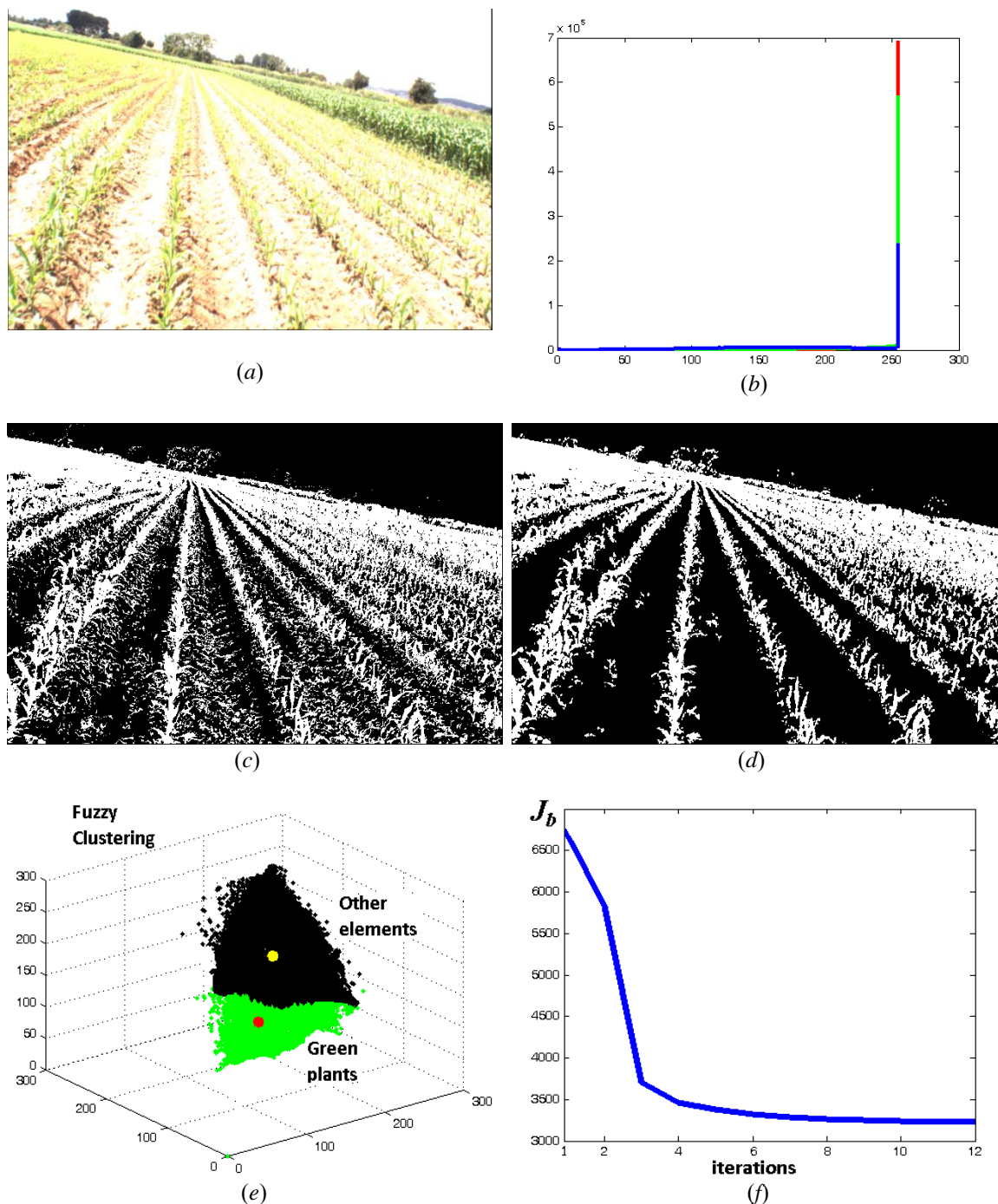
With respect the convergence criterion one can see that the convergence is relatively fast,  $J_b$  drops drastically at the first iterations, iteration two at *figures 5.9(f)* and *5.10(f)* and three at *figure 11(f)*. Although this convergence process is only applied for images classified as with insufficient contrast, we can infer again that the number of iterations is relatively low. This confirms what we had already advanced with regard the maximum number of iterations set to  $t_{max} = 14$ , which validates the setting of this parameter.



**Figure 5.9:** (a) Original image representative of the set of images with sufficient contrast; (b) its histogram for the three RGB spectral channels; (c) binary image obtained with COM; (d) binary image obtained with FC; (e) distribution of clusters and centers obtained through FC; (f) variation of the criterion function against the number of iterations.



**Figure 5.10:** (a) Original image representative of the set of images with insufficient contrast; (b) its histogram for the three RGB spectral channels; (c) binary image obtained with COM; (d) binary image obtained with FC; (e) distribution of clusters and centers obtained through FC; (f) variation of the criterion function against the number of iterations.



**Figure 5.11:** (a) Original image representative of the set of images with insufficient contrast; (b) its histogram for the three RGB spectral channels; (c) binary image obtained with COM; (d) binary image obtained with FC; (e) distribution of clusters and centers obtained through FC; (f) variation of the criterion function against the number of iterations.

## 5.6. Conclusions

We propose a new automatic ES for greenness image identification in agricultural images in maize and barley fields. Its open architecture allows its application to any type of crops in growth states where green plants appear. The

proposed system is also applicable to other environments where the main goal is the greenness detection, by example urban areas where trees or lawn densities are of interest.

The proposed ES is based on two main modules, where the first one decides about the image quality by histogram analysis. Depending on this decision, each image is conveniently processed either by applying a combination of classical greenness strategies or through a module where the *FC* approach supports the adjusting of a threshold which is dynamic for each image. The proposed strategy has been tested with a broad set of images coming from different camera devices, all oriented toward automatic applications in agricultural images for site specific treatments and captured under very different environmental conditions.

The ES has been designed with an open architecture, so that in the future be possible to replace or add new modules, being of particular interest to study new automatic thresholding methods (*Avci and Avci 2009*) or improve the decision rule defined in equation (5.3) by applying a decision-tree inductive learning strategy such as ID3, CART or C4.5 among others (*Duda et al. 2000*). In this work we have applied the *DWT* to the three RGB spectral channels; the down-sampling process could be only applied to the intensity image after the image transformation from the RGB color space to HSI (H-hue, S-saturation, I-intensity) among others.

In addition, once green plants are identified, the remainder parts, mainly the soil, could be analyzed to identify relevant ecologically categories (*Luscier et al. 2006*), thus the proposed ES could be extended to deal with the analysis of soil materials with the feedback of samples belonging to this category.

# AUTOMATIC GREENNESS THRESHOLD DETECTION

## 6.1. Introduction

The present algorithm is an automatic procedure for calculating the threshold to be used in greenness segmentation. We use greenness segmentation to delete from the original image those pixels not belonging to vegetation, i.e.: non-green pixels. Depending on the threshold, pixels whose green spectral component is below it are deleted and those whose spectral component is above it remain in the image. Therefore the election of a proper threshold is a critical step on greenness segmentation. A value of the threshold below the correct value would make that pixels not belonging to vegetation remain in the segmented image. On the other hand, a value of the threshold above the correct value would delete pixels belonging to vegetation.

There are two reasons for developing this algorithm:

1. The algorithm is part of the RHEA project (*RHEA 2014*) which consists on a tractor moving along maize crop lines detecting weeds. This implies that the whole process takes part in outdoor conditions. Due to this fact, images coming from the camera are subject to environmental conditions, that is, variation in luminance, brightness, IR and UV radiation, shadows, etc. The value of the threshold for segmentation depends directly on these factors and therefore must be constantly adjusted according to the image received.
2. Since we are working on real time, we cannot stop the whole process to adjust the value of the threshold when necessary. An automatic algorithm for this

adjustment is necessary to run the whole process without interruptions and to warranty the quality of the segmentation.

Thus, every single image is segmented with its proper threshold which is calculated using this algorithm. Then the next image is again segmented with its new threshold and so on.

The whole process used to detect, locate and quantify weeds in RHEA (*RHEA 2014*) consists on the following steps:

1. Image acquisition.
2. Exposure time adjustment for the next image.
3. Quality control of the image.
4. Image segmentation (greenness extraction).
5. Crop lines detection.
6. Weeds percentage calculation.

The algorithm for the automatic greenness threshold calculation belongs to *step 4, image segmentation*. It could be used for any color extraction. In case of a pure color, i.e.: a color with only one of the three RGB spectral components different to zero, the algorithm is only applied to the non-zero spectral component. Otherwise, for multispectral colors, three different thresholds should be calculated and applied to the image, one for each RGB spectral component. Since this algorithm has been developed within RHEA project (*RHEA 2014*), we are only interested on the green spectral component of the image.

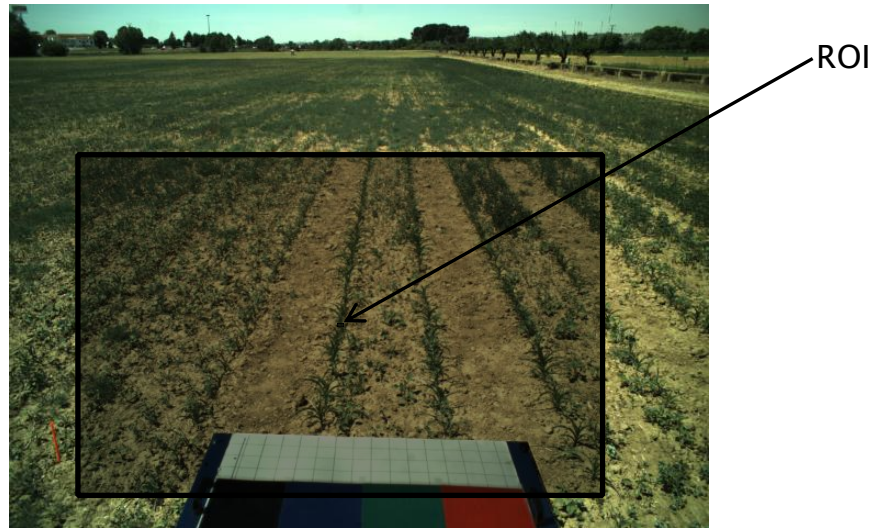
## **6.2. Description of the method**

Once the image is acquired and classified as suitable to be processed, a region of interest, ROI is established to work with, *figure 6.1*. There are two reasons to establish this ROI:

- Despite the pitch angle of the camera pointing downwards to the crops, the field of view of the camera is wide enough to cover on the horizon trees, buildings, sky,

clouds and any other elements apart from crops that may interfere and ruin the results.

- Due to the movements of the tractor while it goes along the crops, a security margin is necessary to compensate these movements of the tractor and also the camera to avoid processing areas of no interest.



**Figure 6.1:** On every image a region of interest ROI is established to ensure that only crop areas are processed and to compensate tractor movements. The rest of the image out of the ROI is not processed.

This method is achieved in four steps which are described as follows:

1. We calculate the percentage of green of every pixel inside the ROI. To do so we use the RGB spectral components obtaining a matrix  $P$  of the same dimensions as the ROI whose values are calculated as follows:

$$P_{ij} = \frac{R_{ij}}{R_{ij} + G_{ij} + B_{ij}} \quad (6.1)$$

where  $P$  is the percentage of green and  $i, j$  are the coordinates (row and column) of the pixel inside the ROI.

2. From every column of  $P$  we calculate the mean of the percentages  $m_j$ :

$$m_j = \frac{1}{f} \sum_{i=1}^f P_{ij} \quad j = 1, \dots, c \quad (6.2)$$

where  $m_j$  is the mean of the percentages of the column  $j$ ,  $f$  is the total number of rows and  $c$  the total number of columns of the ROI.

3. Now we obtain the maximum and minimum values of the mean calculated in *step 3*:

$$Max = \max(m_j);$$

$$Min = \min(m_j);$$

4. Finally we define the value of the threshold to be applied in the image segmentation as the arithmetic mean between the maximum and the minimum:

$$Threshold, T = \frac{Max + Min}{2} \quad (6.3)$$

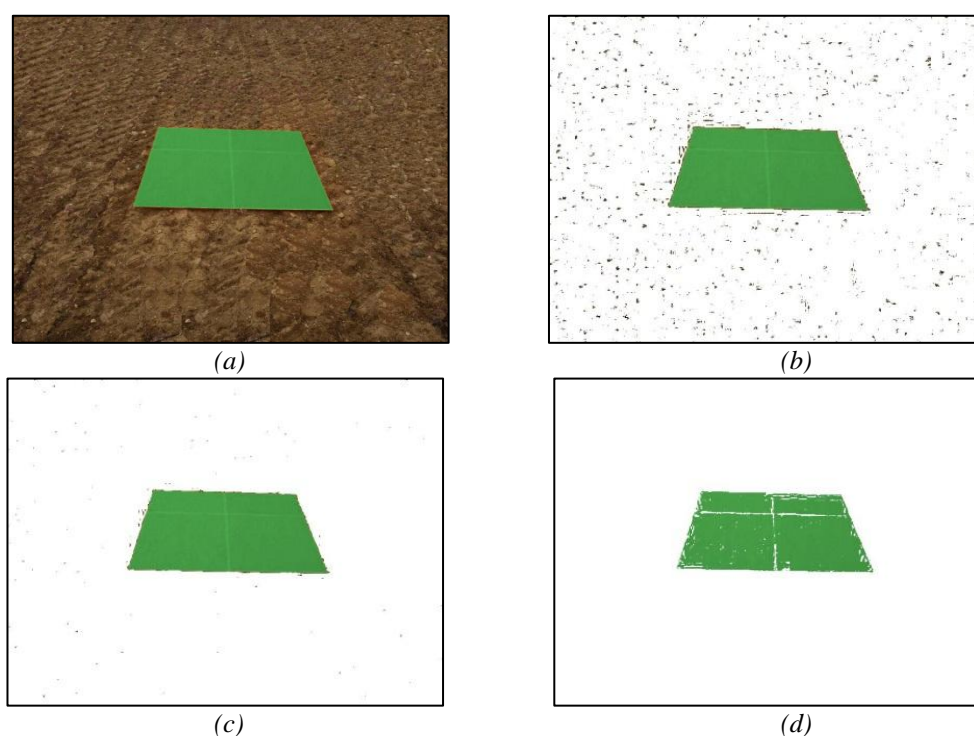
This threshold is applied to the image and those pixels below this value are deleted from the original image remaining only those pixels whose green percentage is above this threshold. Final result is a binarized image where the entire number of pixels present in the image belong to vegetation of any kind, either crops or weeds.

### 6.3. Results

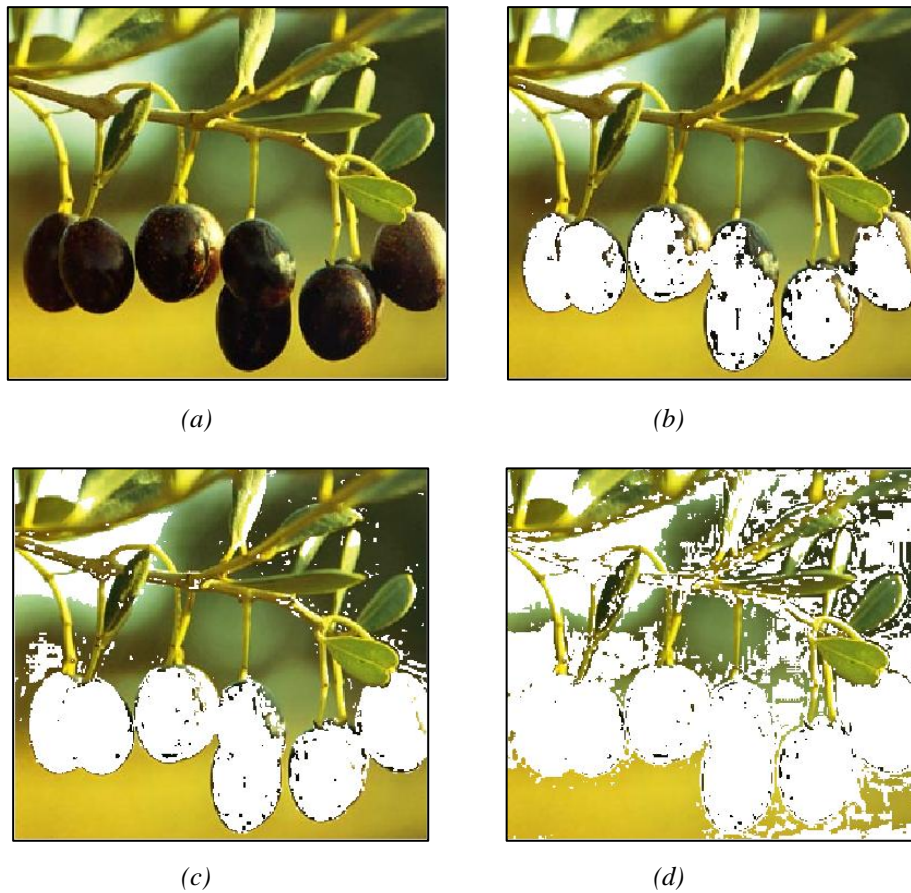
This algorithm has been tested with more than 300 static images from maize crops as well as video images at an approximate rate of 2 images per second. Images were taken in the CSIC-CAR research crop fields of La Poveda, Arganda del Rey, Madrid in different days during May/June in years 2012 and 2013, covering different atmospheric conditions reaching up to 4000 images.

The goodness of the results of this algorithm has been validated by visual inspection since there not exist neither a ground truth nor an unique correct value for the threshold. A threshold value is acceptable if it extracts green pixels properly and if by increasing or decreasing the value in two or three percentage points the results worsens considerably by deleting green pixels or remaining non-green pixels from the original image. Therefore we must be aware that greenness segmentation has not a unique value for the threshold to apply though the range of acceptable values is quite narrow. Usually the acceptable values for a good segmentation cover from 3 to 5 correlative percentage points.

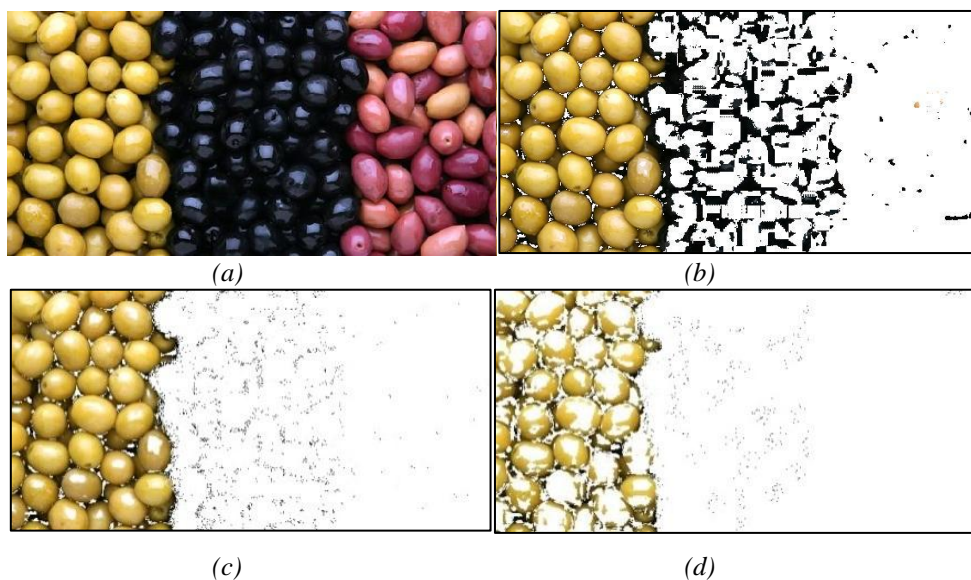
In the following figures different examples of greenness segmentation using this algorithm for calculating the value of the threshold are shown. Images labeled with (a) are the original images. Images labeled with (c) are the resulting images after segmentation using the value of the threshold calculated by our algorithm. Images labeled with (b) are the same images segmented with a threshold 2 or 3 points below the value calculated by our algorithm and finally images labeled with (d) are the images segmented with a threshold 2 or 3 points above the value given by our algorithm. In all the samples we can notice how the value calculated by our algorithm (images labeled with (c) ) is inside the range of appropriate values. However, if we segment the images with a threshold value 2 or 3 points below our value (images labeled with (c) ) pixels not belonging to vegetation begin to remain in the segmented image. On the same way, if we increase the value of the threshold 2 or 3 points above our value, green pixels begin to disappear in the segmented image.



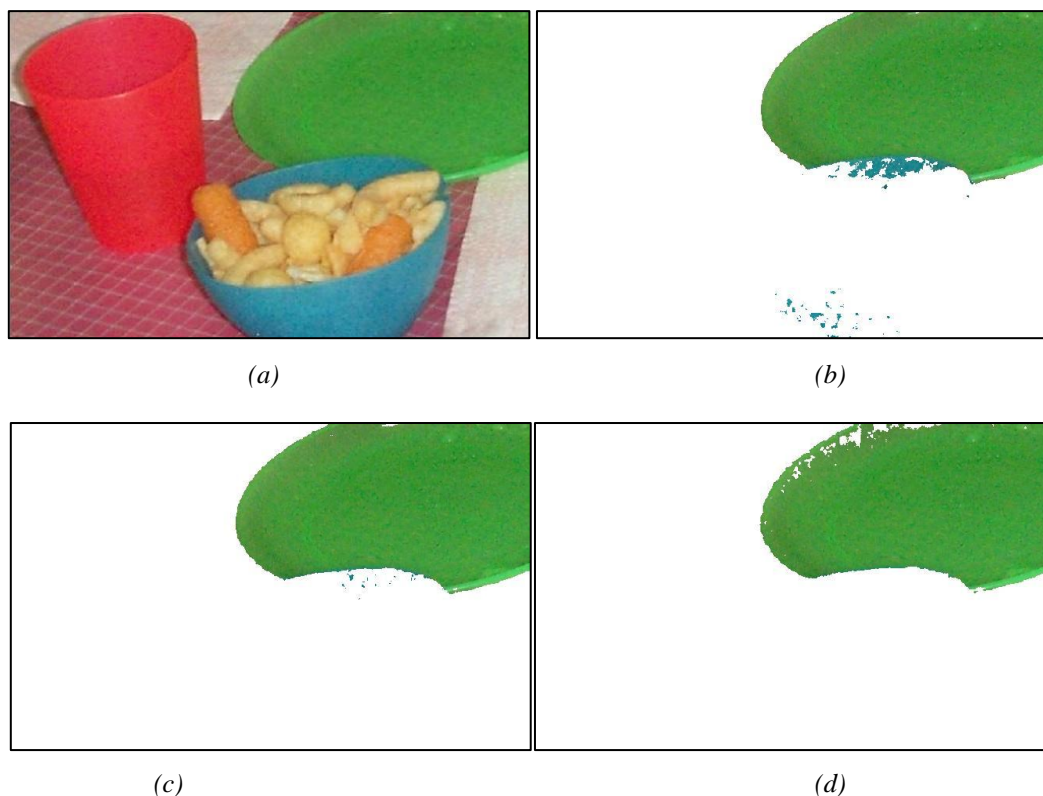
**Figure 6.2:** (a) One meter square cardboard, Green colored placed on unplanted soil. (b), (c) and (d) show the segmented image applying a threshold value of 36%, 38% and 40% respectively. (c) Shows the image segmented using the value given by our algorithm which in this case is 38%. Below this value, i.e. applying a threshold of 36% instead of 38%, non-green pixels begin to appear in the segmented image (b). On the same way applying a threshold of 40% instead of 38%, green pixels begin to disappear in the segmented image (d).



**Figure 6.3:** (a) black olives over a dark background (b), (c) and (d) show the segmented image applying a threshold value of 36%, 39% and 42% respectively. (c) Shows the image segmented using the value given by our algorithm which in this case is 39%. Below this value, i.e. applying a threshold of 36% instead of 39%, non-green pixels begin to appear in the segmented image (b). On the same way applying a threshold of 42% instead of 38%, green pixels begin to disappear in the segmented image (d).



**Figure 6.4:** (a) Original image of green, black and purple olives (b), (c) and (d) show the segmented image applying a threshold value of 36%, 38% and 40% respectively. (c) Shows the image segmented using the value given by our algorithm which in this case is 38%. Below this value, i.e. applying a threshold of 36% instead of 38%, non-green pixels begin to appear in the segmented image (b). On the same way applying a threshold of 40% instead of 38%, green pixels begin to disappear in the segmented image (d).



**Figure 6.5:** (a) Different colored objects (b), (c) and (d) show the segmented image applying a threshold value of 42%, 45% and 48% respectively. (c) Shows the image segmented using the value given by our algorithm which in this case is 45%. Below this value, i.e. applying a threshold of 42% instead of 45%, non-green pixels begin to appear in the segmented image (b). On the same way applying a threshold of 48% instead of 45%, green pixels begin to disappear in the segmented image (d).

## 6.4. Conclusions

A new algorithm has been proposed for calculating the value of the threshold to apply in the image segmentation whose results are within the acceptable range of values. Though there no exist a ground truth or any other way to verify the results, visual inspection is the method used to validate the goodness of the algorithm. The range of acceptable values is very narrow, usually not more than three to five correlative percentage points. Using this algorithm, the value of the threshold is always within this range of values. Approximately 300 static images and 4000 video images have been analyzed and in all cases the value of the threshold has been within the range of acceptable values.

The formulas used are as simple as means. However, satisfactory results have also been obtained using other mathematic operations such as geometric mean, variance, etc.

A proposal for future works is to develop a data base of crop images, in the same way as the *Middlebury (2014)* stereoscopic data base images (*Scharstein and Pal 2007*), so that we could have a ground truth to compare the results of applying different algorithms for image segmentation.

## REFERENCIAS / REFERENCES

1. AGL (2008). *PERcepción vIsual, toma de deCIsión y Actuación en sistemas artificiales complejos: control selectivo de malas hierbas en cultivos agrícolas (PERICIA)*. AGL2008-04670-C03-02/AGR Plan Nacional de I+D+i, Ministerio de Educación y Ciencia, España.
2. AGL (2011). **Sistema de Inspección Terrestre en Vehículos autónomos y su Aplicación efectiva a la Detección de Malas hierbas y su Control Localizado (GroW)** (AGL2011-30442-C02-02). Plan Nacional de I+D+i, Ministerio de Economía y Competitividad, España.
3. Agrawal S., Halder A., Biswas S., Sarkar D. and Sarkar, P.P. (2007). *A novel method for estimation of distance of an object from camera*. IET-UK International Conference on Information Communication Technology in Electrical Sciences (ICTES), pp. 751-753.
4. Asif M., Amir S., Israr A. and Faraz M. (2010). *A vision system for autonomous weed detection robot*. International Journal of Computer and Electrical Engineering, vol. 2, no. 3, pp. 486–491.
5. Astrand B. and Baerveldt A.J. (2005). *A vision based row-following system for agricultural field machinery*. Mechatronics, vol. 15, no. 2, pp. 251–269.
6. Astrand B. (2005). *Vision based perception or mechatronic weed control*. Ph.D. thesis, Chalmers and Halmstad Universities, Sweden, Stockholm.
7. Astrand B. and Baerveldt A.J. (2002). *An agricultural mobile robot with vision-based perception for mechanical weed control*. Auton. Robots 2002, 13, 21–35.
8. Avci E. and Avci D. (2009). *An expert system based on fuzzy entropy for automatic threshold selection in image process*. Expert Systems with Applications, 36(2), 3077–3085.
9. Aziz M., Mertsching B. (2010). *Fast depth saliency from stereo for region-based artificial visual attention*. Lecture Notes in Computer Science, 6474, 367-378.

10. Balasko B., Abonyi J. and Feil B. (2008). *Fuzzy Clustering and Data Analysis Toolbox for Use with Matlab*, Veszprem University, Hungary, Budapest.
11. Baret F., de Solan B. Lopez-Lozano R., Ma K. and Weiss M. (2010). *GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5° zenith angle: Theoretical considerations based on 3D architecture models and application to wheat crops*. Agric. For. Meteorol., 150, 1393–1401.
12. Bezdek J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, NY, USA.
13. Billingsley J. and Schoenfisch M. (1997). *The successful development of a vision guidance system for agriculture*” Computers and Electronics in Agriculture, vol. 16, no. 2, pp. 147–163.
14. Billingsley J. and Schoenfisch M. (1995). *Vision-guidance of agricultural vehicles*. Auton. Robots, 2, 65–76.
15. Birchfield S. and Tomasi, C. (1998). *Depth discontinuities by pixel-to-pixel stereo*. Sixth International Conference on Computer Vision, pp. 1073-1080.
16. Bossu J., Gée Ch., G. Jones and Truchetet F. (2009). *Wavelet transform to discriminate between crop and weed in perspective agronomic images*. Computers and Electronics in Agriculture, vol. 65, no. 1, pp. 133–143.
17. Bossu J., Gée Ch., Guillemin J. P. and Truchetet F. (2006). *Development of methods based on double Hough transform and Gabor filtering to discriminate crop and weeds in agronomic images*. Proc of the SPIE 18th Annual Symposium Electronic Imaging Science and Technology, vol. 6070, no. 23, San Jose, Calif, USA, 2006.
18. Buch N., Orwell J., Velastin S. (2009). *3D Extended Histogram of Oriented Gradients (3DHOG) for Classification of Road Users in Urban Scenes*. Proc. of BMVC, September 2009.
19. Burgos-Artizzu X. P., Ribeiro A., Guijarro M. and Pajares, G. (2011). *Real-time image processing for crop/weed discrimination in maize fields*. Computers and Electronics in Agriculture, 75, 337–346, (2011).

20. Burgos-Artizzu X.P., Ribeiro A., Tellaeche A., Pajares G., Fernández-Quintanilla C. (2009). *Improving weed pressure assessment using digital images from an experience-based reasoning approach*. Computer Electronic Agriculture, 65, 176–185.
21. Cantzler H. (2013). *An overview of range images*. Disponible: [http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\\_COPIES/CANTZLER2/range.html](http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/CANTZLER2/range.html) (accedido Enero 2014).
22. Chen Z., Pears N. and Liang, B. (2006). *A method of visual metrology from uncalibrated images*. Pattern Recognition Letters, 27, 1447-1456.
23. Cohen N. A *Color Balancing Algorithm for Cameras*. Available online: [http://www.stanford.edu/class/ee368/Project\\_11/Reports/Cohen\\_A\\_New\\_Color\\_Balancing\\_Method.pdf](http://www.stanford.edu/class/ee368/Project_11/Reports/Cohen_A_New_Color_Balancing_Method.pdf) (accedido Enero 2014).
24. Criminisi A., Reid I. and Zisserman A. (2000). *Single view metrology*. International Journal of Computer Vision, 40 (2), 123-148.
25. Cyberbotics. Webots: Robot Simulator, (2014). disponible: <http://www.cyberbotics.com/> (accedido Enero 2014).
26. Daolei W. and Lim, K. (2011). *Obtaining depth map from segmented-based stereo matching using graph cuts*. Journal of Visual Communication and Image Representation, 22, 325-331.
27. Davies G., Casady W., Massey R. (2013). *Precision agriculture: an introduction*. Water Quality Focus Guide (WQ450, disponible <http://extension.missouri.edu/explorepdf/envqual/wq0450.pdf>) (accedido Enero 2014).
28. DPI (2009). **Sistema de Vigilancia, Búsqueda y Rescate mediante colaboración de vehículos marinos y aéreos** (DPI2009-14552-C02-01). Ministerio de ciencia e Innovación, España.
29. Duda R. O., Hart P. E. and Stork D. S. (2000). *Pattern Classification*. John Wiley & Sons, New York, NY, USA.
30. Faugeras, O.D., Toscani, G. (1986). *The calibration problem for stereo*. Proc. IEEE Computer Vision and Pattern Recognition, pp. 15–20.
31. Fernandes J. C. and Neves J. A. B. (2006). *Angle invariance for distance measurement using a single camera*. IEEE International Symposium on Industrial Electronics, pp. 676-680.

32. Fontaine V. and Crowe T. G. (2006). *Development of line-detection algorithms for local positioning in densely seeded crops*. Canadian Biosystems Engineering, 48, 7.19–7.29.
33. Fu K.S., Gonzalez R.C. and Lee C.S.G. (1988). *Robótica: Control, Detección, Visión e Inteligencia*. McGraw Hill: Madrid, Spain, pp. 306–336.
34. Gebhardt S. and Kauhbauch W. A. (2007). *A new algorithm for automatic Rumex obtusifolius detection in digital image using colour and texture features and the influence of image resolution*. Precision Agriculture, 8(1), 1–13.
35. Gebhardt S., Schellberg J., Lock R. and Kauhbauch, W. (2006). *A. Identification of broad-leaved dock (Rumex obtusifolius L.) on grassland by means of digital image processing*. Precision Agriculture. 7(3), 165–178.
36. Gée Ch., Bossu J., Jones G. and Truchetet F. (2008). *Crop/weed discrimination in perspective agronomic images*. Computers and Electronics in Agriculture, 60(1), 49–59.
37. Gonzalez R. C. and Woods R. E. (2008). *Digital image processing*. Upper Saddle River, NJ: Pearson Prentice Hall.
38. Gonzalez R. C., Woods R. E. and Eddins S. L. (2009). *Digital Image Processing Using MATLAB*, Prentice Hall, Upper Saddle River, NJ, USA.
39. Guerrero J.M., Pajares G., Montalvo M., Romeo J. and Guijarro, M. (2012a). *Support vector machines for crop/weeds identification in maize fields*. Exp. Syst. Appl., 39, 11149–11155.
40. Guerrero J.M., Romeo J., Emmi L., Montalvo M., Guijarro M., Pajares G., Gonzalez-de-Santos P. (2012b). *Influence of the Vision System Pitch Angle on Crop and Weeds Detection Accuracy*. First RHEA International Conference on Robotics and Associated High-Technologies and Equipment for Agriculture, Pisa, Italy, 19–21 September 2012; pp. 319–324.
41. Guerrero J.M., Montalvo M., Romeo J., Oliva D., Guijarro M., Pajares G. (2012c). *Detection of images with high CCD charge level due to indirect sunlight in maize fields*. Int. Conf. of Agricultural Engineering (CIGR-AgEng2012), July 8-12, Valencia (Spain), 6 pages.
42. Guerrero J.M., Guijarro, M., Montalvo M., Romeo J., Emmi, L., Ribeiro, A., Pajares, G. (2013). *Automatic expert system based on images for accuracy crop row detection in maize fields*. Exp. Syst. Appl., 40, 656–664.

43. Guijarro M., Pajares G., Riomoros I., Herrera P.J., Burgos-Artizzu X.P. and Ribeiro A. (2011). *Automatic segmentation of relevant textures in agricultural images*. Comput. Electron. Agric., 75, 75–83.
44. Guijarro M., Guerrero J.M., Montalvo M., Romeo J., Pajares G. (2013). *A new approach to solve image thresholding in precision agriculture*. 9th European Conference on Precision Agriculture. Lleida, Catalonia, Spain, July 7-11, poster. <http://www.ecpa2013.udl.cat/>
45. Hague T., Marchant J. A. and Tillett D. (1997). *A system for plant scale husbandry*. Precision Agriculture, pp. 635–642.
46. Hague T., Tillett N.D. and Wheeler H. (2006) *Automated crop and weed monitoring in widely spaced cereals*. Precis. Agric., 7, 21–32.
47. Hall, E.L., Tio, J.B.K., McPherson, C.A., Sadjadi, F.A. (1982). *Measuring curved surfaces for robot vision*. Computer Journal, 15, 42–54.
48. Holub O. and Ferreira S. T. (2006). *Quantitative histogram analysis of images*. Computer Physics Communications, 175, 620–623.
49. Horn B. (1986). *Robot Vision*, MIT Press.
50. Hough. C. (1962). *A method and means for recognizing complex patterns*. U.S. Patent Office No. 3069654.
51. Hsu C., Lu M., Wang W. and Lu Y. (2009). *Distance measurement based on pixel variation of CCD images*. ISA Transactions, 48, 389-395.
52. ISCAR (2014). Ingeniería de Sistemas Automática y Robótica. <http://www.dacya.ucm.es/area-isa/> (accedido Enero 2014).
53. Ito, M. (1991). *Robot vision modelling-camera, modelling and camera calibration*. Adv. Robotics 5, 321–335.
54. Jeon H.Y., Tian L.F. and Zhu H. (2011). *Robust crop and weed segmentation under uncontrolled outdoor illumination*. Sensors, 11, 6270–6283.
55. Ji R. and Qi L. (2011). *Crop-row detection algorithm based on Random Hough Transformation*. Mathematical and Computer Modelling, 54(3-4), 1016–1020.
56. Jiang G.Q., Zhao, C.J. and Si Y.S. (2010). *A Machine Vision Based Crop Rows Detection for Agricultural Robots*. Proc. of the 2010 International Conference on Wavelet Analysis and Pattern Recognition, Qingdao, China, 11–14 July 2010; pp. 114–118.

57. Jinlin X. and Liming X. (2010). *Autonomous Agricultural Robot and Its Row Guidance*. Proc. of the 2010 International Conference on Measuring Technology and Mechatronics Automation, Nanjing, China, 13–14 March 2010; Volume 1, pp. 725–729.
58. Jones G., Gée Ch. and Truchetet F. (2009a) *Assessment of an interrow weed infestation rate on simulated agronomic images*. Computers and Electronics in Agriculture, 67(1-2), 43–50.
59. Jones G., Gée Ch. and Truchetet F. (2009b) *Modelling agronomic images for weed detection and comparison of crop/weed discrimination algorithm performance*. Precision Agriculture, 10(1), 1–15.
60. Kataoka T., Kaneko T., Okamoto H. and Hata S. (2003). *Crop growth estimation system using machine vision*. Proc. 2003 IEEE/ASME international conference on advanced intelligent mechatronics, Kobe, Japan, July 20 -24, 2003.
61. Kaushik K. T. (2013) *Formulation of a n-degree polynomial for depth estimation using a single image*. Cornell University Library, (2012). Available [http://arxiv.org/find/cs/1/au:+Tiwari\\_K/0/1/0/all/0/1](http://arxiv.org/find/cs/1/au:+Tiwari_K/0/1/0/all/0/1), (accedido Enero 2014).
62. Kirk K., Andersen H. J., Thomsen A. G. and Jørgensen J. R. (2009). Estimation of leaf area index in cereal crops using red–green images. Biosystems Engineering, 104, 308–317.
63. Kise M. and Zhang Q. (2008). *Development of a stereovision sensing system for 3D crop row structure mapping and tractor guidance*. Biosystems Engineering, 101(2), 191–198.
64. Kise M., Zhang Q. and Rovira-Mas F. (2005). *A stereovision-based crop row detection method for tractor-automated guidance*. Biosystems Engineering, 90(4), 357–367.
65. Krotkov E. (1987). *Focusing*. *International Journal of Computer Vision*, 1(3), 223-237.
66. Kumar-Mada S., Smith M.L. Smith L.N. and Midha S. (2002). *An overview of Passive and Active Vision Techniques for Hand-Held 3D data acquisition*. Optical Metrology, Imaging and Machine Vision (A. Shearer, F.D. Murtagh, J. Mahon, P. F. Whelan, Eds.). Proc. of SPIE, 4877, 16-27.

67. Leemans V. and Destain M. F. (2006). *Application of the Hough transform for seed row localisation using machine vision*, Biosystems Engineering, 94(3), 325–336.
68. Leibe B., Cornelis N. and Cornelis K., Van Gool L. (2007). *Dynamic 3D Scene Analysis from a Moving Vehicle*. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'07), Madison, Wisconsin, June 17-22, pp. 1–8.
69. Li M., Imou K., Wakabayashi K. and Yokoyama S. (2009). *Review of research on agricultural vehicle autonomous guidance*. Int. J. Agric. Biol. Eng., 2, 1–26.
70. Ling P. P. and Ruzhitsky V. N. (1996). *Machine vision techniques for measuring the canopy of tomato seedling*. Journal Agricultural Engineering Research, 65(2), 85–95.
71. Lobay A. and Forsyth D.A. (2006). *Shape from texture without boundaries*. Int. J. Computer Vision, 67(1), 71–91.
72. López-Granados, F. (2011). *Weed detection for site-specific weed management: Mapping and real-time approaches*. Weed Res., 51, 1–11.
73. Luscier J. D., Thompson W. L., Wilson J. M., Gorham B. E. and Dragut L. D. (2006). *Using digital photographs and object-based image analysis to estimate percent ground cover in vegetation plots*. Frontiers in Ecology and the Environment, 4(8), 408–413.
74. Marchant J. (1996). *Tracking of row structure in three crops using image analysis*. Computers and Electronics in Agriculture, 15(2), 161–179.
75. Meyer G. E. and Camargo-Neto J. (2008). *Verification of color vegetation indices for automated crop imaging applications*. Computers and Electronics in Agriculture, 63, 282–293.
76. Meyer G. E., Camargo-Neto J., Jones D. D. and Hindman T. W. (2004). *Intensified fuzzy clusters for classifying plant, soil, and residue regions of interest from color images*. Computers and Electronics in Agriculture, 42, 161–180.
77. Meyer, G.E., Hindman, T.W. and Lakshmi, K. (1999). **Machine vision detection parameters for plant species identification**. In G.E. Meyer and J.A. DeShazer (Eds.) Precision Agriculture and Biological Quality. Proceedings of SPIE, Bellingham, WA 3543:327-335.

78. Middlebury (2014). Middlebury Stereo Datasets. Available online: <http://vision.middlebury.edu/stereo/data/> (accedido Enero 2014).
79. Montalvo M., Guerrero J.M., Romeo J., Herrera P.J., Guijarro M., Ribeiro A. (2011). *Camera System geometry for site specific treatment in precision agriculture*. In Proc. First RHEA Workshop (G. Rabatel and P. González de Santos), September 9, Montpellier (France). 4 pages.
80. Montalvo M., Pajares G., Guerrero J.M., Romeo J., Guijarro M., Ribeiro A., Ruz J.J. and de la Cruz J.M. (2012a). *Automatic detection of crop rows in maize fields with high weeds pressure*. Exp. Syst. Appl. 39, 11889–11897.
81. Montalvo M., Guerrero J.M., Romeo J., Oliva D., Guijarro M., Pajares G. (2012b). *Unsupervised learning for crop/weeds discrimination in maize fields with high weed densities*. Int. Conf. of Agricultural Engineering (CIGR-AgEng2012), 5 pages, July 8-12, Valencia (Spain), (<http://cigr.ageng2012.org>).
82. Montalvo M., Guerrero J.M., Romeo J., Emmi, L., Guijarro M., Pajares G. (2013a). *Automatic expert system for weeds/crops identification in images from maize fields*. Exp. Syst. Appl. 40, 75–82.
83. Montalvo M., Guerrero J.M., Romeo J., Guijarro M., De la Cruz J.M., Pajares G. (2013b). *Acquisition of Agronomic Images with Sufficient Quality by Automatic Exposure Time Control and Histogram Matching*. Advanced Concepts for Intelligent Vision Systems (ACIVS'13) (J. Blanc-Talon, A. Kasinski, W. Philips, D. Popescu, P. Scheunders, Eds.), Lecture Notes in Computer Science, vol. 8192, pp. 37-48, Springer International Publishing Switzerland. October 28-31, Poznan, Poland.
84. Moreno-Matías, P., Gómez-Gil, J. (2007). **Automatic guidance of a tractor using computer vision**. Pattern Recognition and Machine Intelligence. Lecture Notes in Computer Science Volume 4815, 2007, pp 169-176
85. National Instruments (2014). LabView. Available online: <http://spain.ni.com/> (accedido Enero 2014).
86. Neto J. C. (2004). *A combined statistical–soft computing approach for classification and mapping weed species in minimum tillage systems*. Lincoln, NE: University of Nebraska.
87. Olsen H. J. (1995). *Determination of row position in small-grain crops by analysis of video images*. Computers and Electronics in Agriculture, 12(2), 147–162.

88. Onyango C. M. and Marchant J. A. (2003). *Segmentation of row crop plants from weeds using colour and morphology*. Computers and Electronics in Agriculture, 39(3), 141–155.
89. Otsu N. (1979). *A threshold selection method from gray-level histogram*. IEEE Trans. System Man and Cybernetics, 9, 62–66.
90. Ozden K.E., Schindler K. and van Gool L. (2010). *Multibody Structure-from-motion in practice*. IEEE Trans. Pattern Analysis and Machine Intelligence, 32(6), 1134-1141.
91. Pajares G. and de la Cruz J. M. (2004). *A wavelet-based image fusion tutorial*. Pattern Recognition, 37, 1855–1872.
92. Payet N. and Todorovic S. (2011). *Scene Shape from Texture of Objects*. Proc. IEEE Computer Vision and Pattern Recognition (CVPR'11), Colorado Springs, CO, pp.2017-2024.
93. Pla F., Sanchiz J. M., Marchant J. A. and Brivot R. (1997). *Building perspective models to guide a row crop navigation vehicle*. Image and Vision Computing, 15(6), 465–473.
94. Prados E. and Faugeras O. (2006). *Shape from shading*. Handbook of mathematical models in computer vision, Springer pp. 375-388.
95. Reid J.F. and Searcy S.W. (1997). *Vision-based guidance of an agricultural tractor*. IEEE Control Syst. 7, 39–43.
96. RHEA (2014). *Robot Fleets for Highly Effective Agriculture and Forestry Management*. Disponible: <http://www.rhea-project.eu/> (accedido Enero 2014).
97. Ribeiro A., Fernández-Quintanilla C., Barroso J. and García-Alegre M. C. (2005). *Development of an image analysis system for estimation of weed*. Proc. 5th European Conference on Precision Agriculture (ECPA '05), vol. 1, no. 1, pp. 169–174.
98. Romeo J., Pajares G., Montalvo M., Guerrero J.M., Guijarro M., Ribeiro A. (2012a). *Crop row detection in maize fields inspired on the human visual perception*. The ScientificWorld Journal, vol. 2012, Article ID 484390, 10 pages.
99. Romeo J., Oliva D., Guerrero J.M., Montalvo M., Guijarro M., Pajares G. (2012b). *Plants discrimination in maize fields applying human perception to digital images*. Proc. Int. Conf. of Agricultural Engineering (CIGR-

- AgEng2012), July 8-12, Valencia (España), 6 pages, ISBN-10: 84-615-9928-4 / ISBN-13: 978-84-615-9928-8 (<http://cigr.ageng2012.org>).
100. Romeo J., Guerrero J., Montalvo M., Emmi L., Guijarro M., Gonzalez-de-Santos P., Pajares G. (2013a). *Camera Sensor Arrangement for Crop/Weed Detection Accuracy in Agronomic Images*. Sensors 2013, 13, 4348-4366.
  101. Romeo J., Pajares G., Montalvo M., Guerrero J.M., Guijarro M., de la Cruz J.M. (2013b). *A new expert system for greenness identification in agricultural images*. Exp. Syst. Appl. 2013, 40, 2275–2286.
  102. Rosin P.L. and Ioannidis E. (2003). *Evaluation of global image thresholding for change detection*. Pattern Recognition. Letters, 24, 2345–2356.
  103. Rovira-Mas, F. (2011). **Global 3D Terrain Maps for Agricultural Applications**. Advances in Theory and Applications of Stereo Vision, Dr. Asim Bhatti (Ed.), InTech.
  104. Rovira-Mas, F., Zhang, Q., Hansen, A.C. (2010). **Mechatronics and Intelligent Systems for Off-road Vehicles**, Springer-Verlag.
  105. Rovira-Mas F., Zhang Q. and Reid J. F. (2008). *Stereo vision three-dimensional terrain maps for precision agriculture*. Computers and Electronics in Agriculture, 60(2), 133–143.
  106. Rovira-Mas F., Zhang Q., Reid J. F. and Will J. D. (2005). *Hough transform-based vision algorithm for crop row detection*. Journal Automobile Engineering, Part D, vol. 219(8), 999–1010.
  107. Rovira-Mas F., Zhang Q., Reid J.F. and Will J.D. (2003). *Machine vision based automated tractor guidance*. Int. J. Smart Eng. Syst. Des., 5, 467–480.
  108. Ruiz-Ruiz G., Gómez-Gil J. and Navas-Gracia L.M. (2009). *Testing different color spaces based on hue for the environmentally adaptive segmentation algorithm (EASA)*. Computers and Electronics in Agriculture, 68, 88–96.
  109. Sainz-Costa N., Ribeiro A., Burgos-Artizzu X., Guijarro M. and Pajares G. (2011). *Mapping wide row crops with video sequences acquired from a tractor moving at treatment speed*, Sensors, 11(7), 7095–7109.
  110. Salvi, J., Armangu, X., Batlle, J. (2002). **A comparative review of camera calibrating methods with accuracy evaluation**. Pattern Recognition, 35, 1617-1635.

111. Saxena A., Chung S.H. and Ng A.Y. (2008). ***3-D Depth Reconstruction from a Single Still Image***. Int. J. Computer Vision, 76, 53-69.
112. Scharstein D. and Szeliski R. (2002). ***A taxonomy and evaluation of dense two-frame stereo correspondence algorithms***. International Journal of Computer Vision, 47(1/2/3), 7–42.
113. Scharstein D. and Pal C. (2007). ***Learning conditional random fields for stereo***. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 07), Minneapolis, MN, June 2007.
114. Schneider Kreuznach (2013a). ***C-Mount Lenses compact series I***. Available online: <http://www.schneiderkreuznach.com/en/industrial-solutions/lenses-and-accessories/products/> (accedido Enero 2014).
115. Schneider Kreuznach (2013b). ***Tipps and Tricks***. (Available online: [http://www.schneiderkreuznach.com/tipps/uv-ir\\_cut\\_filter.htm](http://www.schneiderkreuznach.com/tipps/uv-ir_cut_filter.htm), (accedido Enero 2014).
116. Shan-Shan C., Wu-heng Z. and Zhi-lin F. (2011). ***Depth estimation via stereo vision using Burchfield's algorithm***. Proc. IEEE 3rd International Conference on Communication Software and Networks (ICCSN), Beijing, China, May 27-29, pp. 403-407.
117. Shrestha D. S., Steward B. L. and Birrell S. J. (2004). ***Video processing for early stage maize plant detection***. Biosystems Engineering, 89(2), 119–129.
118. Slaughter D.C, Giles D.K and Downey D. (2008). ***Autonomous robotic weed control systems: a review***. Computers and Electronics in Agriculture, 61(1), 63–78.
119. Søgaard H. T. and Olsen H. J. (2003). ***Determination of crop rows by image analysis without segmentation***. Computers and Electronics in Agriculture, 38(2), 141–158.
120. Strand B. A and Baerveldt A. J. (2005). ***A vision based row-following system for agricultural field machinery***. Mechatronics, 15(2), 251–269.
121. Subbarao M. and Surya G. (1994). ***Depth from defocus: A spatial domain approach***. International Journal of Computer Vision, 13(3), 271-294.
122. Subbarao M., Choi T. and Nikzad A. (1993). ***Focusing techniques***. Optical Engineering, 32(11), 2824-2836.

123. SVS-VISTEK (2014). *The Focal Point of Machine Vision*, disponible online: <http://www.svsvistek.com/> (accedido Enero 2014).
124. Tao A., Gang J. and Jian W. (2010). *Auto-focusing Technique for Moving Object Based on Image Definition Criterion*. Proc. 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), vol. 2, pp. 37-42.
125. Tellaeche A., Burgos-Artizzu X.P., Pajares G., Ribeiro A. (2008a). *A new vision-based approach to differential spraying in precision agriculture*. Comput. Electron. Agric., 60, 144–155.
126. Tellaeche A., Burgos-Artizzu X.P., Pajares G., Ribeiro A. (2008b). *A vision-based method for weeds identification through the Bayesian decision theory*. Pattern Recognition, 41, 521–530.
127. Tellaeche A., Pajares G., Burgos-Artizzu X. P., and Ribeiro A. (2011). *A computer vision approach for weeds identification through Support Vector Machines*. Applied Soft Computing, 11(1), 908–915.
128. The Matworks (2014), <http://www.mathworks.com/> (accedido Enero 2014).
129. Tian L.F. and Slaughter D.C. (1998). *Environmentally adaptive segmentation algorithm for outdoor image segmentation*. Computers and Electronics in Agriculture, 21, 153–168.
130. Toshev A., Makadia A. and Daniilidis K. (2009). *Shape-based object recognition in videos using 3D synthetic object models*. Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'09), Miami, Florida, USA, June 20-25, pp. 288-295.
131. Trucco E. and Verri A. (1998). *Introductory Techniques for 3D Computer Vision*. Prentice Hall, Upper Saddle River, NJ.
132. Tsai. R. (1987). *A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses*. IEEE Journal of Robotics and Automation, 3(4), 323–344.
133. Vioix J.B., Douzals J.P., Truchetet F., Assémat L. and Guillemain J.P. (2004). *Spatial and spectral method for weeds detection and localization*. Eurasip Journal on Advances in Signal Processing, 7, 679–685.
134. Wang G., Hu Z., Wu F. and Tsui H. (2005). *Single view metrology from scene constraints*. Image and Vision Computing, 23, 831-840.

135. Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M. (1999). *Shape from Shading: A Survey*. IEEE Trans. Pattern Anal. Image Processing, 21(8), 690-706.
136. Wobbecke D.M., Meyer G.E., Von Bargen K. and Mortensen D.A. (1995). *Shape features for identifying young weeds using image analysis*. Trans. Am. Soc. Agric. Eng. 38 (1), 271–281.
137. Zheng L., Shi D. and Zhang J. (2010). *Segmentation of green vegetation of crop canopy images based on mean shift and Fisher linear discriminate*. Pattern Recognition Letters, 31(9), 920–925.
138. Zheng L., Zhang J. and Wang Q. (2009). *Mean-shift-based color segmentation of images containing green vegetation*. Computer Electronics. Agriculture, 65, 93–98.
139. Zimmermann H. J. (1991). *Fuzzy Set Theory and its Applications*. Kluwer Academic, Norwell, Mass, USA.