

**DESARROLLO DE UN SISTEMA  
INTELIGENTE REMOTO PARA EL  
ANÁLISIS DEL CONTENIDO  
POLÍNICO DE LA ATMÓSFERA**

Estela Díaz López

Septiembre, 2014





MASTER UNIVERSITARIO EN INTELIGENCIA ARTIFICIAL AVANZADA:  
FUNDAMENTOS, MÉTODOS Y APLICACIONES

**DESARROLLO DE UN SISTEMA  
INTELIGENTE REMOTO PARA EL  
ANÁLISIS DEL CONTENIDO  
POLÍNICO DE LA ATMÓSFERA**

REALIZADO POR:

**Dña. Estela Díaz López.**

**Ingeniero de Telecomunicación.**

DIRIGIDO POR:

**Dr. D. Mariano Rincón Zamorano.**

**Profesor Titular de Universidad en dpto. de Inteligencia Artificial.  
UNED.**

CO-DIRIGIDO POR:

**Rosa Pérez Badía.**

**Profesor Titular de Universidad en dpto. de Ciencias Ambientales.  
UCLM.**

Madrid, Septiembre de 2014





---

## Agradecimientos

El trabajo presentado en esta memoria ha sido posible gracias a la colaboración de mi director, el Dr. D. Mariano Rincón Zamorano con la Dra. Dña. Rosa Pérez Badía de la UCLM, en un proyecto sobre el análisis polínico de Castilla-La Mancha. Gracias a esta colaboración he conocido el mundo de los granos de polen, unas partículas invisibles que nos rodean y que son muy importantes para mejorar la calidad de vida de las personas.

Gracias a mi director, Dr. D. Mariano Rincón Zamorano, que me ha sabido aconsejar y guiarme en el campo de la investigación permitiendo explorar diferentes áreas que me han permitido conocer el mundo de los granos de polen y desarrollar el proyecto que aquí se presenta. Gracias por tu sinceridad, por tus consejos, tus correcciones y tu paciencia.

También tengo que dar las gracias a la Dra. Dña. Rosa Pérez Badía por gestionar el equipo de la UCLM y enseñarme el campo de análisis del contenido polínico de la atmósfera. Muchas gracias por tus comentarios y correcciones de esta memoria.

Tengo que agradecer a los tres expertos palinólogos, Consolación Vaquero del Pino, Ana Rapp Benito y Jesús Rojo Úbeda, su dedicación para anotar las diferentes muestras utilizadas en este proyecto y sus explicaciones para desarrollar el sistema propuesto.

Por su puesto que tengo que agradecer a mis compañeros de la UNED las horas que hemos pasado juntos, todo lo que he aprendido de ellos y con ellos. Gracias Tomas, Amparo, Santiago, Alba, Antonio, Dani, Manuel, ....

Gracias a mi familia, a mis padres y a mi hermana, que me ha apoyado en todo momento en mi decisión de dedicarme a la investigación a pesar de que no sean buenos tiempos para ello.

Y sobre todo muchas gracias a Dani, que ha demostrado una gran fuerza y vitalidad en este año tan duro que hemos pasado, apoyandome en todo momento, viendo siempre el lado bueno de las cosas y siendo siempre tan sincero.

A todos muchas gracias.



## RESUMEN

La tarea de detección y recuento polínico es un proceso complejo que necesita de expertos entrenados para diferenciar los granos de polen de otras partículas en muestras procedentes de la succión de aire. Es un proceso necesario, que permite informar del contenido polínico de la atmósfera. El recuento polínico se realiza actualmente de forma manual, mediante microscopía óptica y es muy costoso, tanto temporal como físicamente.

Para mejorar la labor del experto palinólogo y mejorar la información actual sobre la distribución de los granos de polen se propone la utilización de un sistema que permita detectar e identificar el contenido polínico de muestras procedentes de succión de 24 horas de aire. El sistema desarrollado permite digitalizar toda la muestra a baja resolución con un plano de enfoque variable. A partir de la imagen obtenida los expertos pueden anotar sobre las imágenes la posición y el tipo del grano de polen detectado. Si la identificación no es posible a baja resolución, el sistema permite adquirir imágenes 3D de alta resolución centradas en el punto anotado para su identificación. Por si no fuera posible la identificación en esta segunda fase, el sistema desarrollado permite la conexión remota con el microscopio para capturar nuevas imágenes que permitan la identificación del grano.

El sistema desarrollado permite analizar los granos de polen anotados manualmente y además, se está mejorando mediante el desarrollo de módulos que tratan de imitar la labor del experto y que realizan una detección e identificación automática de la imagen, lo que permitirá en un futuro al experto palinólogo solo realizar tareas de verificación de los resultados del sistema automático.

Se han desarrollado los módulos de digitalización y detección a baja resolución, y los resultados obtenidos son prometedores.

**PALABRAS CLAVE:** detección, identificación, grano de polen, aerotrasportado, automatización, digitalización, segmentación

## ABSTRACT

Detection and identification of pollen grains in real air samples is a complex task that requires trained experts to distinguish pollen grains from other particles. It is a necessary process that informs of the pollen grains content of the atmosphere. Nowadays, the counting of pollen grains is done manually with light microscopy. It is a time consuming and hard task.

To improve the palynologist work and the information of the distribution of pollen grains has been developed a computational system to detect and identify the pollen content of 24 hours of air caption samples. The system allow the total digitalization of the sample with autofocus plane and low resolution. The experts are able to detect and identify manually pollen grains with the developed system. If the identification is not possible with a low resolution image, the system can acquire 3D high resolution images centered on a possible pollen grain. Also, there are a third step to identify the object, with the remote connection with the microscope to acquire new 3D high resolution images.

The developed system analyzed the manual points of pollen grains and furthermore, it is being improved by new modules that try to imitate the expert labor. The automatic system, that has been developing, analyze the image to detect and identify pollen grains. In the future, the labor expert only will be the verification of the automatic system results.

At the moment, digitalization and detection module has been developed, with promising results.

**KEYWORDS:**detection, identification, pollen grain, the airborne, automation, digitization, segmentation





# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Identificación del problema . . . . .	1
1.2. Motivación . . . . .	3
1.3. Objetivos . . . . .	3
1.4. Metodología . . . . .	4
1.5. Organización del documento . . . . .	5
<b>2. Antecedentes</b>	<b>7</b>
2.1. Redes de Aerobiología . . . . .	7
2.2. Monitorización polínica aerotrasportada . . . . .	8
2.3. Morfología polínica . . . . .	12
2.4. Métodos de detección y conteo . . . . .	14
2.4.1. Método manual . . . . .	14
2.4.2. Métodos automáticos . . . . .	15
2.5. Resumen del capítulo . . . . .	18
<b>3. Caracterización de los granos de polen</b>	<b>19</b>
3.1. Detección . . . . .	19
3.2. Delineación a partir de puntos semilla . . . . .	21
3.3. Análisis de distribución polínica anotada manualmente . . . . .	23
3.4. Análisis de características visuales . . . . .	24
3.4.1. Color . . . . .	26
3.4.2. Textura . . . . .	28
3.4.3. Forma y tamaño . . . . .	33
3.5. Resumen del capítulo . . . . .	35
<b>4. Sistema inteligente remoto de digitalización, detección e identificación de granos de polen</b>	<b>37</b>
4.1. Introducción . . . . .	37
4.2. Sistema a baja resolución . . . . .	39

4.2.1. Módulo de digitalización . . . . .	39
4.2.2. Modulo de detección . . . . .	41
4.2.3. Módulo de identificación . . . . .	46
4.3. Sistema a alta resolución . . . . .	47
4.3.1. Módulo de digitalización . . . . .	47
4.3.2. Módulo de detección . . . . .	47
4.3.3. Módulo de identificación . . . . .	48
4.4. Sistema de verificación . . . . .	49
4.5. Resumen del capítulo . . . . .	49
<b>5. Resultados experimentales</b>	<b>51</b>
5.1. Materiales . . . . .	51
5.2. Métodos . . . . .	51
5.2.1. Digitalización . . . . .	53
5.2.2. Detección . . . . .	55
5.2.3. Identificación . . . . .	62
5.3. Resumen del capítulo . . . . .	62
<b>6. Conclusiones y trabajo futuro</b>	<b>65</b>
6.1. Discusión . . . . .	65
6.2. Conclusiones . . . . .	66
6.3. Aportaciones a la investigación . . . . .	67
6.4. Trabajo futuro . . . . .	68
<b>A. Protocolo de digitalización y análisis</b>	<b>69</b>
A.1. Digitalización de la muestra . . . . .	69
A.2. Análisis de la muestra . . . . .	72
A.3. Digitalización de polen no identificado . . . . .	77
A.4. Verificación de puntos semilla anotados . . . . .	78
<b>B. Características analizadas de los granos de polen</b>	<b>81</b>
<b>C. Resultados de los clasificadores analizados</b>	<b>85</b>
C.1. Clasificador con todas las características analizadas . . . . .	85
C.2. Clasificador con el mejor subconjunto de las características analizadas	87
<b>D. Imagen y características de los tipos polínicos identificados</b>	<b>93</b>
<b>Listado de referencias</b>	<b>99</b>



# Índice de figuras

1.1. Estaciones de muestreo españolas . . . . .	2
2.1. Estaciones polínicas de la Red de Aerobiología de Castilla-La Mancha	8
2.2. Captador volumétrico Hirst . . . . .	10
2.3. Diagrama de una preparación de una muestra de 24 h . . . . .	11
2.4. Imagen de varios granos de polen teñidos con fucsina . . . . .	11
2.5. Imagen microscopio óptico vs electrónico . . . . .	12
2.6. Morfología de un grano de polen . . . . .	12
2.7. Tipos morfológicos de un grano de polen . . . . .	13
2.8. Área sub-muestreada según metodología REA, cuatro barridos horizontales continuos y equidistantes entre si y del borde de la preparación . . . . .	15
3.1. Interfaz de usuario del programa de visualización y anotación . . . . .	20
3.2. Pila de imágenes a diferentes enfoques y alta resolución . . . . .	21
3.3. Número de granos detectados e identificados de forma manual . . . . .	23
3.4. Distribución anual de tipos polínicos en el año 2012, con radio en función de la cantidad de granos identificados . . . . .	24
3.5. Número de granos identificados de forma manual por tipo . . . . .	25
3.6. Distribución de tipos polínicos a lo largo del año 2012 por muestra . . . . .	25
3.7. Número de meses de polinización por tipo de grano de polen detectado a lo largo del año 2012 . . . . .	26
3.8. Análisis de color en espacio RGB de los tipos polínicos que aparecen en al menos 10 ocasiones . . . . .	27
3.9. Análisis de color en espacio HSV de los tipos polínicos que aparecen en al menos 10 ocasiones . . . . .	28
3.10. Análisis de color inter-muestra de los granos de polen en HSV . . . . .	29
3.11. Proceso de generación de matriz GLCM con vecindad 1 y ángulo $0^{\circ}$ . . . . .	29
3.12. Disposición de los anillos concéntricos para análisis de textura . . . . .	29
3.13. Análisis de textura en anillos concéntricos: Entropía . . . . .	31

3.14. Análisis de textura en anillos concéntricos: Contraste . . . . .	32
3.15. Anillos defectuosos en el grano de polen tipo Pinus . . . . .	32
3.16. Análisis de textura en anillos concéntricos: Correlación . . . . .	33
3.17. Tamaño de los granos de polen por tipo: Area . . . . .	34
3.18. Tamaño de los granos de polen por tipo: Diámetro eje mayor y menor	34
3.19. Outliers a partir de análisis del tamaño . . . . .	35
4.1. Esquema de funcionamiento del sistema semiautomático de análisis .	38
4.2. Muestra diaria digitalizada a baja resolución . . . . .	39
4.3. Situación de la muestra digitalizada sobre la preparación respecto al origen de coordenadas de la base motorizada . . . . .	40
4.4. Situación de los granos en el eje z de la cinta adhesiva . . . . .	40
4.5. Plano de enfoque que mejor se adapta a la muestra . . . . .	40
4.6. Malla de 49 puntos equiespaciados para analizar enfoque . . . . .	41
4.7. Esquema de funcionamiento del sistema de detección de granos de polen	43
4.8. Ejemplo de objetos muy oscuros presentes en las muestras . . . . .	44
4.9. Diferente pigmentación rosácea de los granos de polen . . . . .	45
4.10. Segmentación mediante Transformada de Hough de granos adyacentes	46
4.11. Resultados de entrenamiento y test de árbol de decisión para identificación polínica a 100 aumentos . . . . .	47
4.12. Pila de imágenes que componen la imagen 3D del grano de polen . .	48
4.13. Granos de polen Plantago, Poaceae y Populus a 400 aumentos . . . .	48
5.1. Fotografía del microscopio óptico y digital motorizado junto al PC con software NIS Elements . . . . .	52
5.2. Módulos de configuración del software NIS Elements Basic Research .	52
5.3. Granos de polen a diferentes valores de enfoque . . . . .	53
5.4. Descripción de función de autoenfoco en software NIS . . . . .	54
5.5. Porcentaje de fondo por muestra . . . . .	55
5.6. Esquema de trabajo del algoritmo de detección inicial . . . . .	58
5.7. Resultados de circularidad en la segmentación con transformada de Hough: Separación de objetos cercanos (izda.) y eliminación de objetos no circulares (dcha.) . . . . .	59
5.8. Resultados de entrenamiento y test de árbol de decisión para identificación polínica a 100 aumentos . . . . .	63
A.1. Interfaz de PolenGUI . . . . .	73

---

B.1. Segmentación a partir de punto semilla y disposición de los anillos concéntricos para análisis de color y textura . . . . .	81
C.1. Resultados entrenamiento muestra balanceada con todas las características . . . . .	87
C.2. Resultados evaluación a toda la muestra con el modelo entrenado con muestra balanceada y todas las características . . . . .	88
C.3. Listado de los mejores atributos . . . . .	89
C.4. Resultados entrenamiento muestra balanceada con un subconjunto de las características . . . . .	90
C.5. Resultados evaluación a toda la muestra con el modelo entrenado con muestra balanceada y un subconjunto de las características . . . . .	91



# Índice de tablas

2.1. Muestreadores aerobiológicos . . . . .	9
2.2. Nomenclatura NPC . . . . .	13
2.3. Categoría de granos de polen según tamaño . . . . .	14
2.4. Resumen de estudios con muestras puras de polen . . . . .	17
2.5. Resumen de estudios con muestras reales de polen . . . . .	17
3.1. Resultados de los tres métodos de detección manuales realizados por los expertos . . . . .	22
5.1. Resultados de detección automática con imagen digitalizada a 100 aumentos . . . . .	57
5.2. Comparativa entre algoritmos de detección automáticos . . . . .	59
5.3. Resultados obtenidos tras analizar la segmentación de granos de polen con transformada de Hough circular por línea . . . . .	61
5.4. Resultados obtenidos tras analizar la segmentación de granos de polen con transformada de Hough circular por muestra . . . . .	61
A.1. Listado de tipos polínicos . . . . .	76
D.1. Imagen y características de los granos de polen identificados . . . . .	93
D.1. Imagen y características de los granos de polen identificados . . . . .	94
D.1. Imagen y características de los granos de polen identificados . . . . .	95
D.1. Imagen y características de los granos de polen identificados . . . . .	96
D.1. Imagen y características de los granos de polen identificados . . . . .	97
D.1. Imagen y características de los granos de polen identificados . . . . .	98



# Nomenclaturas

EAN	European Aeroallergen Network, página 1
GLCM	Gray-Level Co-Occurrence Matrix , página 28
GS	Gold-Standard, página 5
REA	Red Española de Aerobiología, página 1
SEAIC	Sociedad Española de Alergología e Inmunología Clínica, página 7





# Capítulo 1

## Introducción

### 1.1. Identificación del problema

Los granos de polen son las esporas masculinas de las plantas con flores. En su interior desarrollan el tubo polínico y las células sexuales (gamétas), cuya función biológica es alcanzar las gamétas de la parte femenina de la flor y llevar a cabo la fecundación. Son partículas resistentes, de forma y tamaño variado. Contienen proteínas que liberadas a la atmósfera son capaces de producir alergias. Su estudio periódico y sistemático es muy importante y tiene aplicaciones diversas. En el área de la botánica, se utiliza para conocer la floración y la evolución de las plantas. En medicina, para conocer la cantidad de pólenes y esporas que influyen en la salud o en agricultura para la predicción de cosechas de plantas como el olivo, que emiten grandes cantidades de polen a la atmósfera. El recuento de partículas biológicas aerotransportadas permite identificar los granos de pólenes que producen polinosis en cada área geográfica, predecir los picos de polinización y adaptar los tratamientos de los pacientes alérgicos.

A pesar de la publicación de varios proyectos y patentes para la automatización del conteo polínico (Branham and Branham, 2012; Boucher et al., 2002b,a; Landsmeer et al., 2009), a día de hoy, el proceso que se lleva a cabo para el recuento del contenido polínico atmosférico continua siendo manual, mediante microscopía óptica. La Red Española de Aerobiología (REA) (Galán et al., 2007) proporciona una metodología estandarizada de trabajo que permite la comparación de datos con distintos grupos, su difusión y la conexión con redes europeas como “The European Aeroallergen Network” (EAN). La EAN<sup>1</sup> se encarga de mantener actualizada una base de datos europea que reúne la información aerobiológica, generada en Europa, por aproximadamente 600 estaciones de muestreo. En España existen más de 56

---

<sup>1</sup><https://ean.polleninfo.eu/Ean/>



Figura 1.1: Estaciones de muestreo españolas

Fuente: [http://www.uco.es/rea/infor\\_rea/estaciones.htm](http://www.uco.es/rea/infor_rea/estaciones.htm)

estaciones repartidas por toda la geografía, como se puede ver en la figura 1.1.

La tarea de detección e identificación manual del contenido polínico es muy costosa, tanto temporal como físicamente, pero muy necesaria para conocer el nivel de alérgenos en suspensión y de otras partículas dañinas para la población. Los expertos palinólogos analizan una parte de la muestra diaria con un microscopio óptico en busca de granos de polen y almacenan la cantidad y el tipo para actualizar las bases de datos autonómicas, nacionales y europeas como la REA y la EAN.

Gracias a la evolución de los sistemas informáticos, se han incorporado a los microscopios sistemas de captura digital, que permiten escanear y almacenar las muestras de contenido polínico para su posterior análisis computacional. Diferentes grupos de investigación como (Branham and Branham, 2012; France et al., 2000; Holt et al., 2011; Ronneberger, 2007; Boucher et al., 2002a; Holt and Bennett, 2014; Landsmeer et al., 2009) han comenzado a desarrollar soluciones automáticas o semiautomáticas para ayudar a los expertos en la tarea de detección e identificación de granos de polen con resultados prometedores.

El principal problema es la detección de los granos de polen, ya que se trabaja

con muestras procedentes de la succión de aire. Estas muestras contienen múltiples partículas, tanto bióticas como abióticas, que dan lugar a multitud de problemas de solapes, oclusiones, partículas rotas, etc., y que dificulta la definición de un sistema de visión artificial. Una vez localizado el grano, la identificación del tipo se puede realizar utilizando una imagen 3D de mayor resolución centrada en el grano.

## 1.2. Motivación

El avance de la tecnología en las últimas décadas, tanto hardware como software, ha permitido generar grandes cantidades de datos que han de ser explicados por y para la comunidad científica. La digitalización de datos desde etapas tempranas permite el procesamiento y la compartición de la información de manera más sencilla y sistemática. Además, también es posible definir sistemas computacionales que imiten la labor del experto y ahorren tiempo y recursos respecto a un análisis manual, proporcionando unos resultados más robustos e independientes del observador.

En el campo de la aerobiología, como se ha comentado anteriormente, se utiliza un método manual para el análisis de muestras polínicas. Se realizan recuentos manuales de las partículas biológicas mediante microscopía óptica y se obtiene la cantidad y el tipo polínico para actualizar diferentes bases de datos.

Gracias a un proyecto de investigación científica y desarrollo tecnológico llevado a cabo en la Universidad de Castilla-La Mancha, titulado “Análisis del contenido polínico y actividad alérgica de la atmósfera de Castilla-La Mancha” y dirigido por Rosa Pérez Badía, y en el que también ha participado la UNED a través del dpto. de Inteligencia Artificial, hemos conocido el problema del recuento polínico aerotransportado. Esto ha dado pie al desarrollo de un sistema computacional cuyo objetivo es facilitar la labor de los expertos palinólogos. Este sistema permitirá mejorar la metodología actual mediante la utilización de herramientas de digitalización y análisis automático de las muestras, de modo que se ahorrará tiempo a los expertos en el conteo y se podrán dedicar más recursos humanos a investigación.

## 1.3. Objetivos

En este trabajo se describe el desarrollo de un sistema computacional semiautomático remoto para la adquisición, detección e identificación de granos de polen procedentes de muestras obtenidas por succión de aire de la atmósfera. Este sistema facilitará el trabajo a los expertos palinólogos y, además, permitirá

aumentar la información actualmente disponible al analizar de manera sistemática más información asociada a los granos de polen (información temporal, espacial y visual).

Este sistema mejorará el protocolo actual al automatizar parte de la labor del experto, dejando únicamente los casos dudosos al especialista para su identificación. El sistema permitirá obtener las siguientes mejoras:

1. Generar una base de imágenes de muestras polínicas diarias.
2. Recopilar un número mayor de características de los granos detectados: tamaño, forma, posición, intensidad, etc.
3. Reducir la carga de trabajo de experto palinólogo, ya que solo tendrá que identificar los objetos dudosos.
4. Mejorar los procesos de predicción polínica, pues permitirá dedicar más recursos humanos a investigación y generar bases de casos más grandes y completas. Las aplicaciones de estos estudios son de enorme interés, por ejemplo, en medicina para el control de alergias o en agricultura para el control de cosechas.

Además, se desarrollarán herramientas con conexión remota para trabajar sobre la muestra a distancia. De esta forma, como último recurso en caso de que la digitalización no permita reconocer el tipo de grano de polen, el experto podrá observar la preparación sin tener que desplazarse a la ubicación del microscopio (que generalmente se encuentran en salas con bajas temperaturas para favorecer el mantenimiento de las preparaciones).

## 1.4. Metodología

La metodología utilizada pretende desarrollar un sistema de conteo de polen automático siguiendo una estrategia incremental. Esto es, se pretende que el sistema desarrollado sea utilizado desde el principio por los expertos y progresivamente ir mejorando su funcionalidad. Para ello, el objetivo número uno es detectar la mayoría de los granos de polen en un entorno real para que después puedan ser clasificados. La clasificación es el segundo objetivo, pues podemos comenzar con un reconocimiento manual, presentando los granos de polen localizados a un experto, y posteriormente construir un clasificador automático que realice esta tarea.

Además, la notación manual de muestras es una tarea tediosa y costosa, por lo que este enfoque pretende conseguir dos objetivos. Por un lado, facilitar la

labor de los expertos y, por otro, contar con la colaboración de los expertos que anotan actualmente las muestras para conseguir un conjunto de entrenamiento más completo, tanto en número de granos de polen anotados como en variedad de tipos, y poder construir un sistema de reconocimiento más preciso.

En el desarrollo del trabajo se seguirán las directrices marcadas por el método científico. En primer lugar, se estudiarán los antecedentes metodológicos y se analizarán los diferentes trabajos publicados en el campo de la detección automática y semiautomática de granos de polen. A continuación, con la ayuda de los expertos, se planificará un procedimiento para obtener una muestra representativa de casos que permita construir un sistema basado en el conocimiento del dominio de los expertos y en los datos aportados por estos casos de ejemplo.

La adquisición se realizará mediante un microscopio óptico motorizado con un sistema de captura digital, conectado a una estación de trabajo con un software que permite su manejo a distancia. A partir de las imágenes digitalizadas de las muestras polínicas, se detectarán de manera automática los granos de polen mediante un sistema de visión artificial. El sistema de visión se compondrá de distintos módulos para la detección de los objetos presentes en la muestra y para la identificación de sus tipos. Estos módulos se diseñarán mediante aprendizaje supervisado a partir de conocimiento del dominio, procedente tanto de las fuentes bibliográficas consultadas como de los expertos, y de los datos de entrenamiento convenientemente anotados.

La validación del sistema automático se realizará mediante la comparación de los resultados con un gold-standar (GS) generado a partir de los casos anotados por los expertos de forma manual. Para ello, se desarrollarán las aplicaciones necesarias que faciliten el conteo e identificación de granos de polen a partir de la imagen digitalizada de la muestra completa. Con la ayuda de estas aplicaciones, los expertos palinólogos anotarán sobre muestras digitalizadas a baja resolución la situación y el tipo de los granos de polen, siguiendo la metodología propuesta por la REA.

## 1.5. Organización del documento

Esta memoria está formada por seis capítulos. En este primer capítulo, el actual, se ha presentado el problema de la detección e identificación de granos de polen aerotrasportados, la motivación de este trabajo y la metodología utilizada. En el capítulo 2 se presenta una revisión bibliográfica sobre el análisis de granos de polen, desde su caracterización morfológica hasta los métodos de conteo empleados. A continuación, en el capítulo 3, se analizan las características visuales, espaciales y temporales de los granos de polen anotados manualmente por los expertos, lo que nos

permitirá tener una primera aproximación al problema. Después se define el sistema de digitalización, detección e identificación de granos de polen propuesto (capítulo 4) y se presentan en el capítulo 5 los resultados experimentales obtenidos tras su implementación y puesta en servicio en el Instituto de Ciencias Ambientales de la Universidad de Castilla-La Mancha. Finalmente, la discusión final, las conclusiones del trabajo y potenciales futuras líneas de investigación se presentan en el capítulo 6.

# Capítulo 2

## Antecedentes

### 2.1. Redes de Aerobiología

Gracias a la agrupación de profesionales de distintas disciplinas en redes autonómicas, nacionales y europeas, se han desarrollado iniciativas para realizar el análisis del contenido polínico de la atmósfera y almacenar los resultados en bases de datos coordinadas, permitiendo difundir contenidos, mejorar conocimiento y optimizar esfuerzos.

En España existen dos redes nacionales para el análisis del contenido polínico en la atmósfera, la Red Española de Aerobiología<sup>1</sup> (REA) y el Comité de Aerobiología<sup>2</sup> de la Sociedad Española de Alergología e Inmunología Clínica<sup>3</sup> (SEAIC). Las redes nacionales coordinan las distintas redes locales, proporcionando una metodología común, lo que permite comparar y compartir resultados y realizar estudios de mayor amplitud y generalidad. Además, proporcionan la información a la base de datos europea “European Allergy Network”<sup>4</sup> (EAN), con sede en Viena, Austria. Las redes locales, formadas mayoritariamente por las comunidades autónomas, son las encargadas de situar los captadores, preparar los elementos de adquisición y procesar las muestras. En la Red de Aerobiología de Castilla-La Mancha<sup>5</sup> (AEROCAM) se dispone de 11 captadores repartidos por 6 localidades de la región (figura 2.1). En función de la época del año se analizan las muestras, diaria o semanalmente, de forma manual siguiendo las indicaciones de la REA. La base metodológica está orientada a obtener resultados totalmente comparables en todas las localidades estudiadas. Por ello, se ha acordado el uso de captadores basados en el sistema Hirst (Hirst,

---

<sup>1</sup><http://www.uco.es/rea/>

<sup>2</sup><http://www.polenes.com>

<sup>3</sup><http://www.seaic.es/>

<sup>4</sup><https://ean.polleninfo.eu/Ean/>

<sup>5</sup><http://aerocam.uclm.es/red>

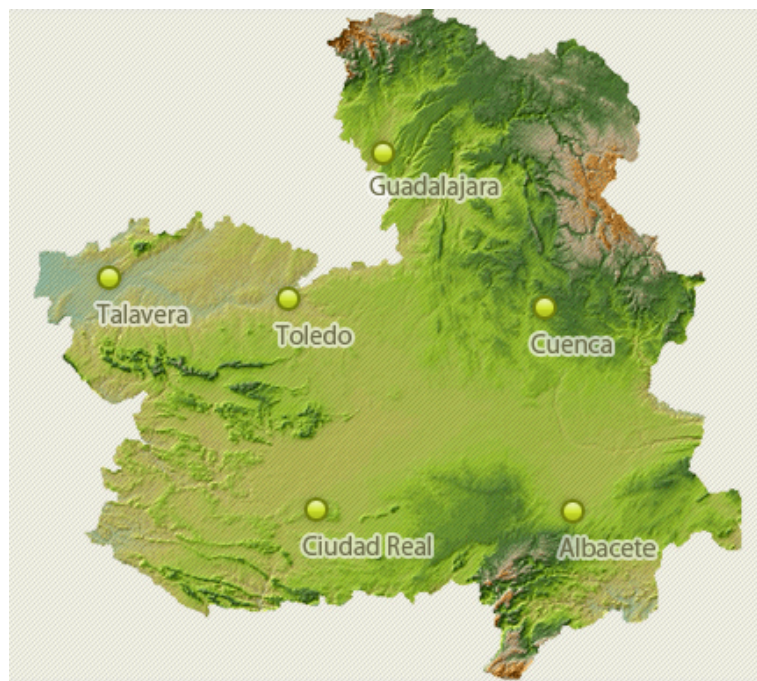


Figura 2.1: Estaciones polínicas de la Red de Aerobiología de Castilla-La Mancha

1952) (modelos comerciales Burkard y Lanzoni), se ha estandarizado el método de muestreo, el sistema de recuento y la obtención de los resultados en todos los centros.

## 2.2. Monitorización polínica aerotrasportada

Desde los primeros estudios aerobiológicos iniciados por Blackely en 1873 (Blackely, 1873) se han desarrollado diferentes sistemas de captación de partículas para muestrear la atmósfera. Los primeros captadores se basaban en la acción de la gravedad para depositar los pólenes (Captador Durham (Durham, 1946)) pero no eran buenos para atrapar esporas y pólenes pequeños. Este tipo de captadores solo permitían cuantificar las partículas por unidad de superficie por lo que para mejorar los resultados, se desarrollaron los captadores por unidad de volumen de aire (Hirst, 1952). Estos muestreadores se basan en el principio de impacto por succión mediante una bomba de vacío, que impulsa el aire aspirado contra la superficie receptora. Otro tipo de captador por impacto son los basados en rotación. Existen múltiples tipos de muestreadores aerobiológicos (ver tabla 2.1) cuya utilización dependerá del objetivo a analizar. En el caso de la monitorización polínica el captador más utilizado es el de tipo Hirst.

El captador desarrollado por Hirst (Hirst, 1952) es un muestreador de impacto por succión. El aparato consiste en una cámara de admisión de aire con una succión



<b>Muestreadores de precipitación</b>	- <b>Precipitación gravimétrica</b>	Placas de Petri Captador Durham Captador Tauber
	- <b>Precipitación electrostática</b>	
	- <b>Precipitación térmica</b>	
<b>Muestreadores de impacto</b>	- <b>Impacto por succión</b>	Captador Hirst
	- <b>Impacto en cascada</b>	Captador Andersen
	- <b>Muestreadores inerciales y ciclónicos</b>	Captador Rotorod Captador ciclónico
<b>Muestreadores de filtración</b>	- <b>Filtros sólidos</b>	Filtros de fibra Captador Cour Filtros por membrana Filtros por cassettes
	- <b>Filtrado en medio líquido</b>	Captador McLeod Multi-Stage Liquid Impinger
<b>Muestreadores biológicos</b>	- <b>Técnicas de biología molecular</b>	
	- <b>Técnicas inmunológicas</b>	

Tabla 2.1: Muestreadores aerobiológicos

Fuente: <http://lap.uab.cat/aerobiologia/es/methods>

de 10 l/min que hace pasar el flujo a través de una hendidura de 14 x 2 mm que está siempre frente a la dirección del viento gracias a que el colector va montado sobre una veleta (fig. 2.2). El aire impacta sobre una superficie receptora compuesta por una cinta adhesiva de 19 mm de ancho y que gira a razón de 2mm/h (48 mm/24h de muestreo). Este sistema hace que las partículas (bióticas, abióticas, minerales y orgánicas) queden retenidas de forma secuencial, permitiendo realizar estudios con precisión horaria.

La colocación de los muestreadores debe ser sobre una superficie horizontal, plana y de fácil acceso. Ha de colocarse en una zona elevada donde no sea tapado por edificios altos, ni próximo a fuentes de emisión masiva de partículas, de forma que se obtengan datos representativos en el radio de cobertura geográfica del muestreador.

Los resultados que se obtienen tras el recuento de la muestra han de ser convertidos a concentraciones medias diarias (pólenes/ $m^3$ ) mediante un simple cálculo matemático en el que intervienen los siguientes factores:

- El número de partículas detectadas (N).
- La superficie de la muestra analizada. Para calcularla se necesita conocer el número (l) de líneas de la preparación sobre las que se realiza el recuento (barridos) y su longitud (K1) y anchura (K2), que dependen del objetivo-ocular que use el microscopio. Con estos parámetros se calcula la proporción

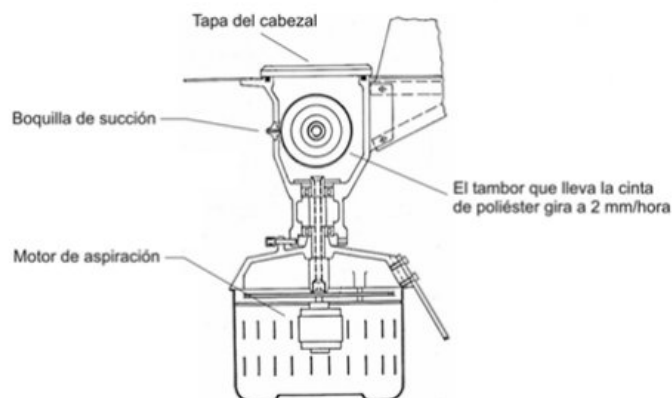


Figura 2.2: Captador volumétrico Hirst

analizada respecto de toda la superficie útil de la preparación microscópica (K3).

- El volumen de aire que aspira el captador, normalmente 10 l/minuto, que al cabo de un día equivalen a  $14,4 \text{ m}^3$  de aire (K4).

Estos parámetros se mantienen constantes mientras no se cambie de microscopio ni de objetivo-ocular, por lo que cálculo matemático queda reducido a una simple multiplicación entre una constante y el número de partículas detectadas en los barridos. La fórmula a aplicar en el caso de la metodología Hirst para obtener la concentración media diaria es  $K \times N$ , con  $K$  igual a la constante calculada con los parámetros K1 a K4.

**Ejemplo de cálculo para un diámetro del campo de visión de 0.45 mm**

Volumen de succión:  $10 \text{ l/min} = 600 \text{ l/hora} = 14400 \text{ l/día} = 14,4 \text{ m}^3$

Diámetro medio del campo de visión al microscopio: 0,45 mm

Área de 1 barrido horizontal =  $48 \text{ mm} \times 0,45 \text{ mm} = 21,6 \text{ mm}^2$

Superficie analizada =  $21,6 \times 4 \text{ barridos} = 86,4 \text{ mm}^2$

Superficie total muestreada =  $48 \text{ mm largo} \times 14 \text{ mm ancho} = 672 \text{ mm}^2$

$K = (672 \text{ mm}^2 / 86,4 \text{ mm}^2) \times (1/14,4)$

$N$  = número de granos de polen en los cuatro barridos.

Contenido de partículas por metro cúbico de aire =  $N \times K = N \times 0,54$

La cinta adhesiva donde se retienen las partículas se divide en fragmentos correspondientes a 24 horas de captación. Cada fragmento es montado en una preparación (fig. 2.3) y teñido con una solución que contiene glicerogelatina teñida con fucsina para facilitar una mejor identificación y recuento de los granos de polen,



Figura 2.3: Diagrama de una preparación de una muestra de 24 h

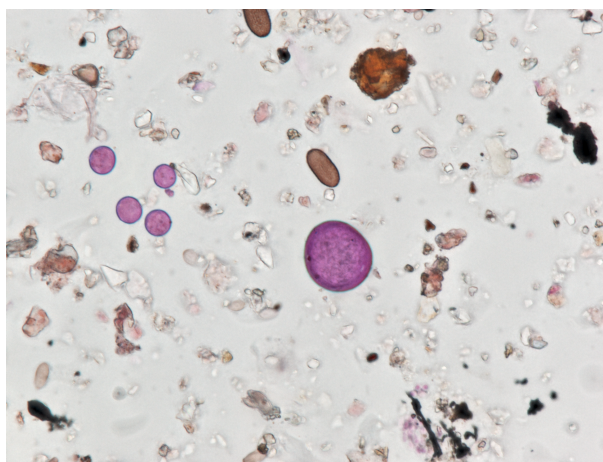


Figura 2.4: Imagen de varios granos de polen teñidos con fucsina

ya que esta tinción es específica para el material vegetal y da una coloración rosada al polen (fig. 2.4).

Existen principalmente dos tipos de microscopios utilizados para el análisis de muestras polínicas, el microscopio óptico y el microscopio electrónico. Cada uno tiene unas ventajas e inconvenientes. En el microscopio óptico la luz atraviesa la muestra y mediante una serie de lentes se obtiene una imagen ampliada de hasta un máximo de 1.000 aumentos (límite debido a que la longitud de onda de la luz visible es de unas 0.6 micras). En el microscopio electrónico se utiliza un haz de electrones a través de un tubo de vacío y permite conseguir aumentos de hasta 1.000.000 de veces. Hay dos tipos de microscopios electrónicos, el de transmisión MET (la muestra es atravesada por los electrones) y de barrido MEB (los electrones rebotan en la superficie de la muestra). Tal como se observa en la figura 2.5, las imágenes proporcionadas por cada tipo de microscopio son muy diferentes.

A pesar de que la calidad de las imágenes es inferior, se utilizan predominantemente los microscopios ópticos en el conteo polínico debido al coste económico y al tiempo para analizar las muestras. Los microscopios ópticos pueden costar entre 100-5000 € mientras que los electrónicos alrededor de 500.000€.

Actualmente, el análisis de las muestras se realiza de forma manual por expertos palinólogos mediante microscopía óptica a 400 aumentos (conseguidos con la

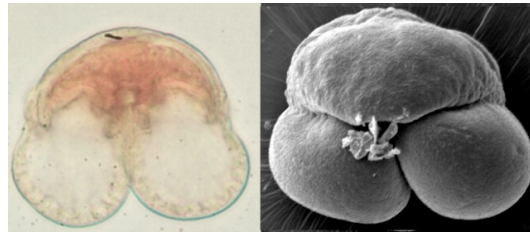


Figura 2.5: Imagen microscopio óptico vs electrónico

Fuente:

<https://sites.google.com/site/microscopiabarrido/Inicio/3—microscopio-electronico-versus-optico>

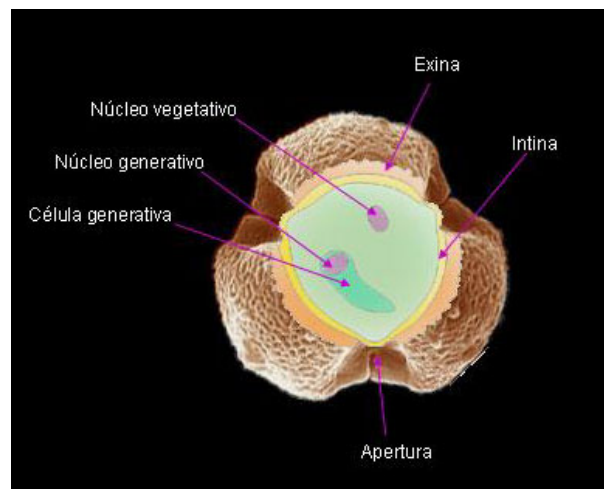


Figura 2.6: Morfología de un grano de polen

Fuente: <http://www.ugr.es/~mcares/Organografia/reproduccion/Gametofito%20conceptos.htm>

combinación de un objetivo de 40x y oculares de 10x) siguiendo las directrices de la REA que se describen en el apartado 2.4.1.

## 2.3. Morfología polínica

El polen maduro presenta una morfología bien definida (ver fig. 2.6) que en general permite identificar la planta de la cual procede. A la hora de describir un grano de polen hay que considerar que se trata de un cuerpo tridimensional que ocupa un volumen en el espacio. Un grano de polen está rodeado de dos cubiertas. La interior se denomina intina, es suave y poco resistente, y rodea al protoplasma. La exterior se denomina exina, es muy resistente y posee poros/aperturas y diferentes tipos de ornamentación (rugosidades).

A nivel internacional existe una nomenclatura que permite aunar en una palabra toda la información necesaria para conocer las aperturas de un grano de polen. Es el sistema NPC, donde N es el número de aperturas, P la posición y C su carácter.

Número (N)	Posición (P)	Carácter (C)
Mono- (1 abert)	-cata- (polo proximal)	-leptomado
Di- (2 abert)	-ana- (polo distal)	-colpado
Tri- (3 abert)	-anacata- (ambos polos)	-porado
Tetra- (4 abert.)	-zono- (una franja ecuatorial)	-colporado
Penta- (5 abert.)	-dizono- (dos franjas ecuatoriales)	-pororado
Hexa- (6 abert.)	-panto- (toda la superficie)	-sulcado
Poli- (muchas abert.)		

Tabla 2.2: Nomenclatura NPC

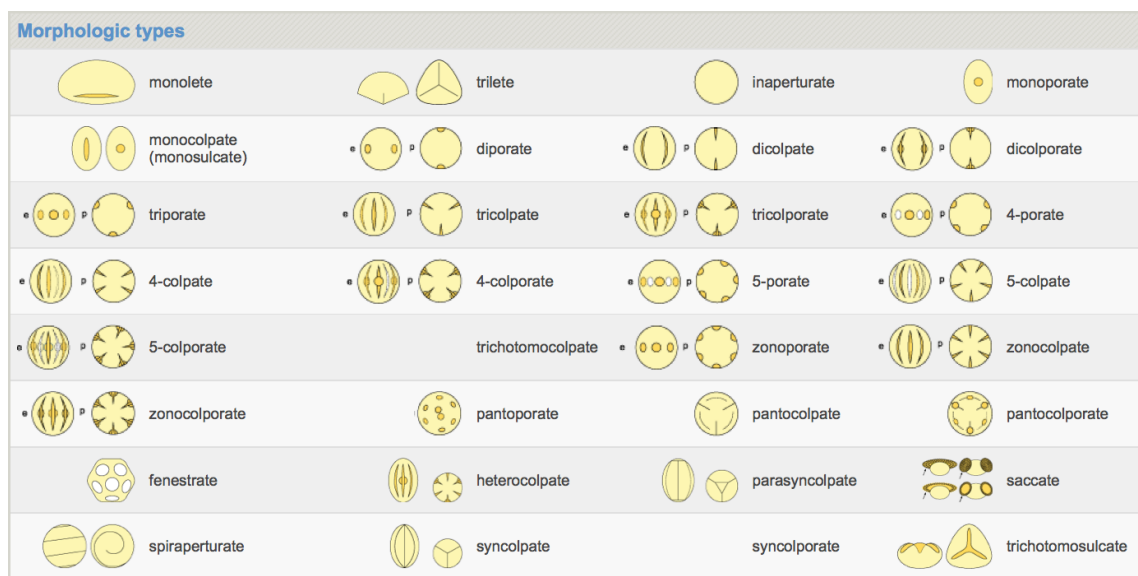


Figura 2.7: Tipos morfológicos de un grano de polen

Fuente: [http://apsa.anu.edu.au/morphologic\\_types](http://apsa.anu.edu.au/morphologic_types)

En la tabla 2.2 se puede ver la nomenclatura establecida. En función del número, la posición y la clase de las aperturas se definen diferentes tipos morfológicos (ver fig. 2.7).

Las características más importantes para los estudios morfológicos de los granos de polen son las siguientes: pared, aperturas, simetría y polaridad, forma, tamaño, superficie y ornamentación.

Uno de los principales problemas a la hora de identificar los tipos de polen es la similitud morfológica al microscopio óptico de diferentes especies. Por ello, se definen los tipos polínicos como un conjunto de caracteres morfológicos que presentan los granos de polen y que son una combinación única que los hace distinguibles e identificables al microscopio. Según (Subiza, 2001) las quenopodiáceas y amarantáceas son imposibles de diferenciar mediante el microscopio óptico y por tanto, los recuentos de ambas familias se tienen que expresar como un único tipo polínico (Chenopo-Amaranthaceae). Otro problema se da al no poder diferenciar

Longitud del eje	Categoría
< 10 $\mu\text{m}$	Muy pequeños
10-25 $\mu\text{m}$	Pequeños
25-50 $\mu\text{m}$	Medianos
50-100 $\mu\text{m}$	Grandes
100-200 $\mu\text{m}$	Muy grandes
200 $\mu\text{m}$	Gigantes

Tabla 2.3: Categoría de granos de polen según tamaño

granos de polen de alergenicidad muy diferente, cuando una misma familia botánica, como las urticáceas, presenta pólenes muy difíciles de identificar a nivel de géneros. Así, uno de sus géneros, *Parietaria*, presenta granos de polen que producen alergias y otro género de la misma familia como *Urtica*, no. Aunque hay ligeras diferencias en el tamaño del grano de polen de estos géneros, tales diferencias no son útiles debido al gran rango de variabilidad intraespecie; por tanto, los recuentos se tienen que expresar conjuntamente como urticáceas.

El tamaño de un grano de polen se determina por la longitud de su eje mayor. Como regla general, se clasifican en función del tamaño en 6 categorías (ver tabla 2.3).

## 2.4. Métodos de detección y conteo

### 2.4.1. Método manual

Actualmente, la detección y recuento de granos de polen obtenidos a partir de muestras procedentes de la succión de aire se realiza de forma manual al microscopio óptico, siguiendo las indicaciones de la REA. El análisis de granos de polen presentes en una muestra diaria requiere de mucho tiempo a los expertos palinólogos. Por ello, solo se realiza un sub-muestreo de, al menos, el 10 % de la preparación (por normativa de la EAN).

Existen tres métodos de sub-muestreo: el método aleatorio, el método longitudinal (horizontal) y el método latitudinal (vertical). El método propuesto por la REA (Galán et al., 2007) es el recuento mediante cuatro barridos horizontales continuos y equiespaciados a lo largo de toda la preparación (fig. 2.8). El número de barridos es importante para realizar una correcta aproximación a la concentración polínica diaria. Según (Sikoparija et al., 2011) un número elevado de barridos requiere mucho tiempo y no aporta grandes mejoras, mientras que un número bajo desvirtúa los resultados.

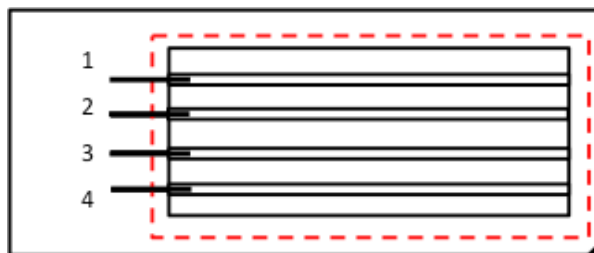


Figura 2.8: Área sub-muestreada según metodología REA, cuatro barridos horizontales continuos y equidistantes entre si y del borde de la preparación

El área sub-muestreada depende del campo de visión de cada microscopio e influye en la conversión entre el número de partículas detectadas y la concentración media. El tiempo necesario por los expertos palinólogos para recorrer los cuatro barridos horizontales depende de la cantidad polínica que contenga la muestra y puede variar entre 15 minutos, para muestras con muy poco contenido polínico, y horas para las muestras de los meses de mayor polinización.

Habitualmente, en el método manual se obtiene la cantidad de granos de cada tipo polínico por día. El análisis por franja horaria también es posible utilizando una plantilla horaria sobre la muestra, pero el incremento de tiempo necesario para la detección y conteo es elevado. Por ello, no existen muchos estudios de evolución polínica horaria (Barnes et al., 2001; Puc, 2012).

### 2.4.2. Métodos automáticos

La automatización del conteo e identificación de granos de polen es un problema complejo que fue propuesto hace más de 40 años por J. R. Flenley (Flenley, 1968). Durante la última década, diferentes grupos de investigación de todo el mundo (Hodgson et al., 2005; Jones, 2012; Rodriguez-Damian et al., 2006; Boucher et al., 2002a; Ranzato et al., 2007) han publicado artículos científicos donde se analiza este problema. De todos los grupos analizados, solo dos trabajan con muestras reales, es decir, procedentes de la captación de polen procedente del aire (Boucher et al., 2002a; Ranzato et al., 2007) mientras que el resto trabaja con preparaciones puras de distintos tipos de granos. Un resumen de diferentes trabajos se presenta en las tablas 2.4 y 2.5, procedente de la tesis doctoral de (Ronneberger, 2007).

La digitalización de la muestra es, en la mayoría de los trabajos, bidimensional, con diferente aumento en función del objetivo de análisis: detección o identificación. Analizan características de forma, color o textura en la fase de segmentación y utilizan diferentes técnicas como redes neuronales, arboles binarios o maquinas vector soporte en la fase de clasificación.

En estudios con muestras puras se obtienen unos porcentajes más altos de reconocimiento. Así, en (Rodríguez-Damian et al., 2006) se analizan características de forma y textura junto a redes neuronales para estudiar 3 tipos diferentes de polen de una misma familia de plantas, las Urticaceae, que tienen características morfológicas muy similares. Se combina la transformada de Hough con modelos de contornos activos (edge, snake y convex-hull) para la detección y delineación de los granos. Se analizan características de forma como el área, el perímetro, etc. junto a descriptores de Fourier y momentos estadísticos. También se analizan diferentes medidas basadas en texturas como estadísticas de primer orden, coeficientes de Haralick o la matriz de co-ocurrencia de niveles de gris. Con todas las características y utilizando una red neuronal como clasificador, consigue una tasa de reconocimiento del 89%. En (Allen, 2006) se alcanza un porcentaje de reconocimiento del 98% mediante un microscopio propio denominado “AutoStage” para un grupo reducido de tipos (siete). En el sistema utilizan 43 características de textura y forma y también redes neuronales para la clasificación. En el caso de muestras reales el porcentaje de reconocimiento es menor, debido a que la muestra contiene otras partículas que dificultan el proceso de análisis. En (France et al., 2000) se analizan muestras preparadas con granos de polen y otros materiales, se utiliza una red neuronal para reconocimiento visual de objetos denominada “PARADISE” (France and Duller, 1999) y se obtiene un porcentaje de reconocimiento del 81.7% con 3 tipos de taxón<sup>6</sup>. En (Boucher et al., 2002a) se analizan muestras con polen procedente del aire teñidas con fucsina, utiliza imagen 2D para la detección de polen e imagen 3D para la identificación mediante el análisis de características de color y forma, obteniendo un 77% de reconocimiento con 30 tipos de polen.

La digitalización de los granos de polen difiere en los diferentes estudios analizados. (Allen, 2006) utiliza para la detección de granos un aumento que produce pixel con un tamaño de 3.8  $\mu\text{m}$ . Con 100 capturas de 640x480 pixeles se digitaliza toda la preparación. En la fase de identificación utiliza un aumento mayor con pixel de 0.415  $\mu\text{m}$  tamaño y se adquieren imágenes de 1024 x 768 pixeles. En (France et al., 2000) se adquieren las imágenes con una cámara CCD monocromática con una resolución de 768 x 576 pixeles capaz de distinguir hasta 256 niveles de gris, con un aumento de 250 magnitudes. Para detección se reduce la imagen a un cuarto de la original para disminuir el tiempo de computo. Se utiliza un aumento de 600 magnitudes en (Boucher et al., 2002a) para digitalizar la muestra a color y se adquieren imágenes 3D de cada grano de polen para identificar su tipo. Para ello se toman 100 imágenes a distintas posiciones del eje Z (vertical) con salto de 0.5

---

<sup>6</sup>Taxón es un grupo de una clasificación científica



Table 1.3: Condensed Overview over other approaches to pollen recognition (**Pure Pollen Samples**)

	sample type	microscopy	segmentation	features	classifier	n samples	n taxa	recogn. rate
<b>Massey University New Zealand (Hodgson et al., 2005)</b>	pure	2D	morph.	texture, shape	neural net.	184	4	96%
(Allen et al., 2006)	"	"	"	"	"	1,400	7	98%
"	"	"	"	"	"	3,800	19	89%
<b>University of Sydney Australia (Jones, 2000)</b>	pure	2D	?	texture, shape	binary trees	16,220	46	76.3%
							80	70.7%
<b>University of Vigo Spain (Rodríguez-Damian et al., 2006)</b>	pure	2D	morph, snake	texture, shape	neural net, SVM	100	3	89%
<b>Delft University of Technology Netherlands (Chen et al., 2006)</b>	pure	3D	thresh.	texture, shape, specialized	"linear normal classifier"	254	3	97.2%

pure: pure pollen samples  
SVM: Support Vector Machine

Tabla 2.4: Resumen de estudios con muestras puras de polen

Fuente: Tesis doctoral de Ronneberger (Ronneberger, 2007)

Table 1.4: Condensed Overview over other approaches to pollen recognition (**Real-world Samples**)

	sample type	microscopy	segmentation	features	classifier	n samples	n taxa	recogn. rate	precision
<b>Bangor University UK (France et al., 2000)</b>	pollen +debris	2D	thresh.	neural net.	neural net.	204	3	81.7%	96.7%
<b>INRIA, Sophia-Antipolis France (Boucher et al., 2002)</b>	air	3D color	morph. +color	shape, color, specialized	Mahalanobis	350	30	77%	?
<b>Caltech and New York University, US (Ranzato et al., 2007)</b>	air	2D	DOG + circle fit	texture	Fisher linear discr. + GMM	3104	8	64,9%	30%

DOG: Difference of Gaussians  
GMM: Gaussian mixture model

Tabla 2.5: Resumen de estudios con muestras reales de polen

Fuente: Tesis doctoral de Ronneberger Ronneberger (2007)

µm.

## 2.5. Resumen del capítulo

Los granos de polen son objetos complejos que llevan siendo analizados desde finales del siglo XIX. Dada su influencia en los humanos, ya sea por temas de salud o por temas agroalimentarios, su recuento es muy importante. Por ello, se han desarrollado redes locales, nacionales y continentales que gestionan la colocación de captadores y realizan recuentos manuales de las muestras para su difusión y análisis.

La evolución de la tecnología y el abaratamiento de los sistemas de almacenamiento de información ha permitido el desarrollo de sistemas computacionales que traten de imitar la labor del experto humano. En la última década se han desarrollado sistemas con resultados prometedores que permitirán en un futuro modificar los protocolos actuales de detección e identificación de granos de polen.

El principal problema se encuentra en la detección de granos de polen cuando se analizan muestras procedentes de la succión de aire. Este tipo de muestras contiene múltiples partículas, algunas de las cuales también presentan características comunes a los granos de polen (como las burbujas), por lo que el objetivo es desarrollar sistemas automáticos de conteo con alta sensibilidad y alta precisión, es decir, que el menor número de falsos positivos y falsos negativos. En la bibliografía consultada no aparecen muchos sistemas de detección de granos en muestras reales ya que los resultados, a día de hoy, no son suficientemente buenos. Una vez que el grano ha sido detectado correctamente, su identificación obtiene buenos resultados cuando se utilizan sistemas de digitalización de alta resolución, que permitan conocer los detalles del grano. Mediante digitalizaciones a diferentes planos de enfoque del grano se consiguen tasas de reconocimiento muy altas, como en (Allen, 2006) del 98% con 7 tipos de taxón. En la mayoría de estudios se utilizan pocos tipos polínicos junto a un conjunto de muestras con un número bajo de granos.

El sistema propuesto en este trabajo pretende identificar un número elevado de tipos polínicos, en concreto se tienen previstos 38 tipos distintos, en muestras procedentes de la atmósfera a partir del análisis de fragmentos de 24 horas de captación. El sistema propuesto se aproxima a los requisitos que debe tener un sistema automático de análisis de granos de polen según (Stillman and Flenley, 1996), que propuso que un buen sistema automático debe de diferenciar al menos 40 tipos polínicos.

## Capítulo 3

# Caracterización de los granos de polen

Como se ha comentado en el capítulo anterior, los granos de polen son partículas bióticas circulares de tamaño variable (10 a 250  $\mu\text{m}$ ) que se encuentran en suspensión en el aire. Su adquisición más común se realiza mediante captadores por succión, de manera que los granos se pegan a una cinta adhesiva que posteriormente se corta y se tiñe con fucsina. Esta tinción proporciona a los granos una coloración rosada que facilita su detección.

El objetivo de este capítulo es conocer las características visuales reales de los granos de polen, generar un conjunto de datos de entrenamiento (GS) a partir de las anotaciones manuales realizadas por los expertos, para desarrollar el sistema automático, y conocer la precisión en la detección e identificación de granos mediante el método manual digital. Para ello, se digitalizan 12 muestras a baja resolución, cada una procedente de un mes del año 2012. Son adquiridas por un mismo captador situado en Toledo y se detectan e identifica de forma manual los posibles granos de polen. Los objetos no identificados se digitalizan a alta resolución de forma tridimensional, es decir, con diferentes planos de enfoque.

### 3.1. Detección

Para la detección de granos de polen se ha desarrollado un programa con una interfaz gráfica sencilla (ver figura 3.1) que permite a los expertos palinólogos visualizar una muestra diaria completa, a baja resolución, y realizar anotaciones manuales sobre la imagen siguiendo las directrices de la REA.

El análisis ideal de los granos de polen debería realizarse con la delineación de cada grano de polen pero, dada la cantidad de granos, el proceso puede ser muy

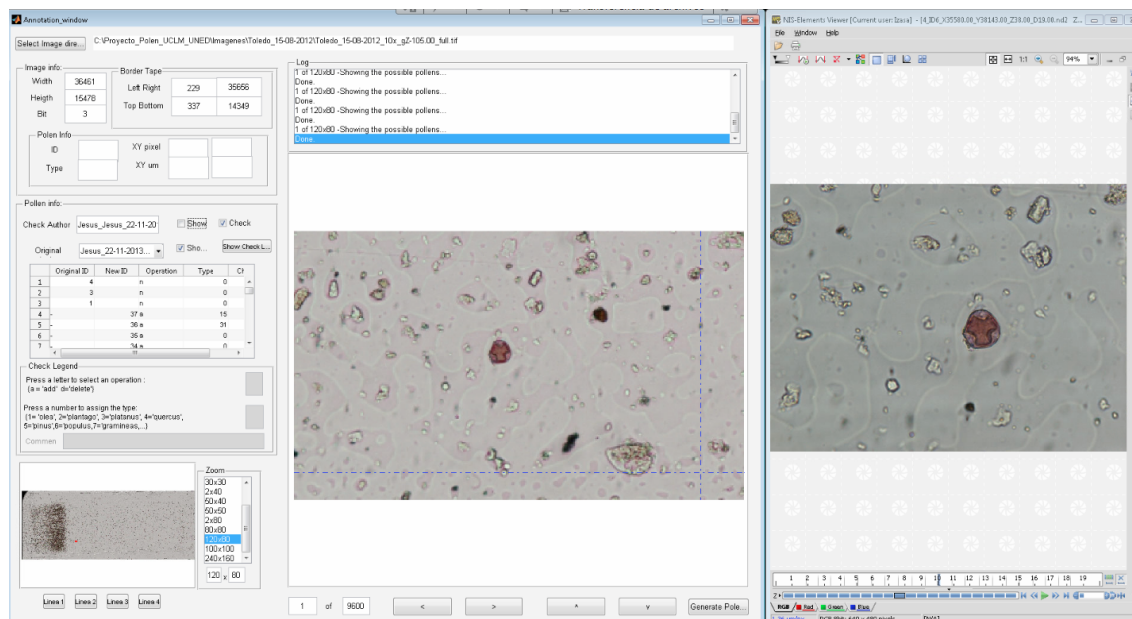


Figura 3.1: Interfaz de usuario del programa de visualización y anotación

lento. Para facilitar el proceso a los expertos, solo se almacena la posición y el tipo de cada grano de polen. A partir del punto anotado se utiliza un algoritmo de segmentación propio (ver pseudocódigo en 3.1) para delinear cada posible grano de polen. En el anexo A.2 se presenta el manual de usuario de este programa, denominado “PolenGUT”.

La detección e identificación ha sido realizada por tres expertos palinólogos. Cada muestra ha sido anotada por al menos un experto y hay muestras anotadas por los tres para analizar la variabilidad inter-experto. Mediante el programa interfaz desarrollado, los expertos han visualizado la preparación a un aumento similar al que utilizan en el microscopio y han recorrido las diferentes líneas. Los granos de polen no reconocidos a baja resolución son digitalizados a una resolución mayor y a diferentes planos de enfoque obteniendo una imagen 3D (ver figura 3.2) que permite su identificación en una segunda fase de visualización.

En la tabla 3.1 se muestra el resumen de los resultados de las tres fases de conteo realizadas de forma manual por los expertos. El método tradicional (óptico) y los métodos digitales a 100 y a 400 aumentos. En total, los expertos han detectado 4061 granos de polen en las 12 muestras analizadas en la fase de detección a baja resolución (100 aumentos), de los cuales 2209 granos (54.4 %) han sido identificados. En la segunda fase de detección a alta resolución (400 aumentos), los expertos han analizado los puntos dudosos, descartando 62 puntos que no eran granos de polen (3.4 %) e identificando 1154 granos (62.3 %). Tras la segunda fase, se tiene un total de 3363 granos de polen identificados y 636 sin identificar. Por tanto, la utilización

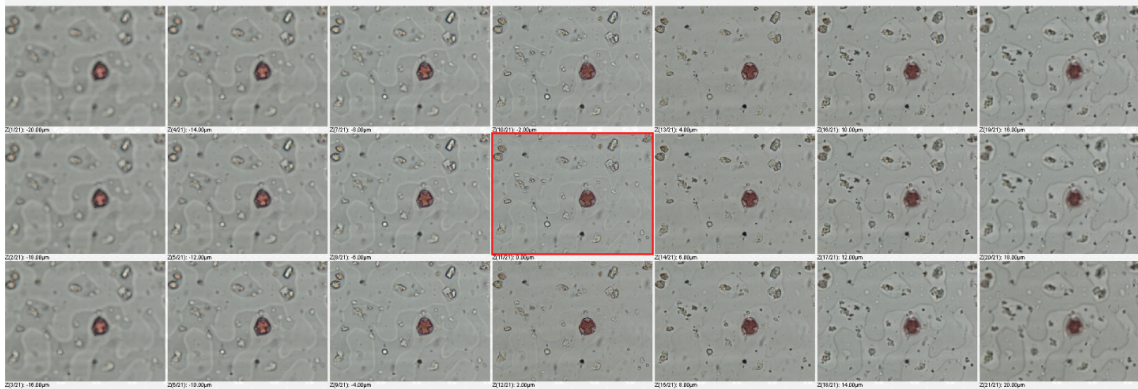


Figura 3.2: Pila de imágenes a diferentes enfoques y alta resolución

de una segunda fase de análisis a mayor resolución ha mejorado la identificación un 28.4 %.

Al final del proceso puede existir un pequeño conjunto de granos de polen sin tipo, a pesar de que se ha realizado el proceso de anotación de forma manual por los expertos. Es debido a dos motivos: a la similitud entre granos de polen y al desenfoque de la imagen utilizada a alta resolución. Hay ocasiones en que la pila de imágenes adquirida a alta resolución no contiene ninguna imagen que permita su identificación, ya sea porque la separación entre imágenes es elevada o porque no se han obtenido imágenes de toda la extensión del grano. Ante duda de tipos, el experto asigna el tipo 0 que corresponde a no identificado. Para estos casos, el sistema desarrollado permite al experto la conexión remota con microscopio para desplazar la platina al grano no identificado en la segunda fase y adquirir nuevas imágenes que permitan su identificación.

Los resultados de los tres métodos de detección manual no se pueden comparar de forma estricta, aunque se puede observar que tienen el mismo orden de magnitud, debido a que que la anchura de la línea analizada en el método digital es ligeramente superior al método tradicional,  $0.5 \mu\text{m}$  frente a  $0.45 \mu\text{m}$ , y a que la es difícil coincidir exactamente en analizar la misma superficie en los barridos horizontales.

### 3.2. Delineación a partir de puntos semilla

A partir de los puntos semilla anotados por los expertos palinólogos, situados en cada grano de polen detectado, se ha desarrollado un algoritmo de segmentación por umbralización en diferentes espacios de color (ver algoritmo 3.1). El algoritmo analiza una imagen centrada en el punto semilla y elimina los objetos oscuros, que corresponden a burbujas o sustancias minerales. También elimina el fondo con un

Estación	Fecha	Detectados			Identificados			No reconocidos		
		Óptico	100x	400x	Óptico	100x	400x	Óptico	100x	400x
Toledo	15/01/12	6	13	13	6	4	4	0	9	9
Toledo	15/02/12	10	26	20	10	10	12	0	16	8
Toledo	15/03/12	1706	1792	1766	1705	809	1741	1	983	25
Toledo	15/04/12	120	161	152	119	98	138	1	63	14
Toledo	15/05/12	1157	1608	1608	1153	1190	1190	4	418	418
Toledo	15/06/12	168	202	202	166	23	139	2	179	63
Toledo	15/07/12	32	44	44	32	5	36	0	39	8
Toledo	15/08/12	7	37	35	7	11	22	0	26	13
Toledo	15/09/12	12	44	31	12	10	20	0	34	11
Toledo	15/10/12	1	51	51	0	4	4	1	47	47
Toledo	15/11/12	0	27	27	0	9	17	0	18	10
Toledo	15/12/12	45	56	50	43	36	40	2	20	10
<b>Total</b>		3264	4061	3999	3253	2209	3363	11	1852	636

Tabla 3.1: Resultados de los tres métodos de detección manuales realizados por los expertos

---

**Algoritmo 3.1** Segmentación de grano de polen a partir de un punto semilla

---

Im= imagen de  $2 \cdot R_{max} \times 2 \cdot R_{max}$  pixeles centrada en punto experto (Seed)

MaskIm1= imagen binaria de pixeles de fondo (umbral variable obtenido del histograma) & pixeles de objetos muy oscuros(umbral fijo de 0.2 analizado empíricamente)

ImColor= Im - MaskIm1

MaskIm2= imagen binaria tras umbralizar Im con el mínimo de ( Seed\_value, umbral Otsu en imagen de  $R_{max} \times R_{max}$  en Seed)

MaskIm3= imagen binaria tras umbralizar ImColor en espacio HSV en función de valor H de Seed

MaskIm4= imagen binaria de círculo de  $R_{min}$  centrado en Seed

PollenSeg= MaskIm4 + ((-MaskIm1) & MaskIm2 & MaskIm3)

---

umbral variable en función del histograma de la imagen, ya que el color blanco del fondo depende de la iluminación de la muestra en la digitalización y de la cantidad de partículas por muestra. Una vez eliminados los pixeles que no son granos de polen, se analiza el resto de la imagen para obtener la delineación del objeto. Para ello se umbraliza la imagen en escala de grises con el mínimo valor obtenido de analizar una subimagen centrada en el punto semilla mediante el método Otsu y el valor del punto semilla. En el espacio de color HSV, que es una transformación no lineal del espacio RGB, el color rojizo se encuentra en valores bajos y altos del canal H. Por ello, en función del valor del canal H del punto semilla se umbraliza el canal H para delinear los objetos rosáceos. Para evitar obtener una segmentación vacía, si en etapas las anteriores no se detecta nada, se utiliza como objeto mínimo un círculo de  $10 \mu m$  de diámetro (tamaño del grano de polen más pequeño esperado) centrado en el punto semilla.

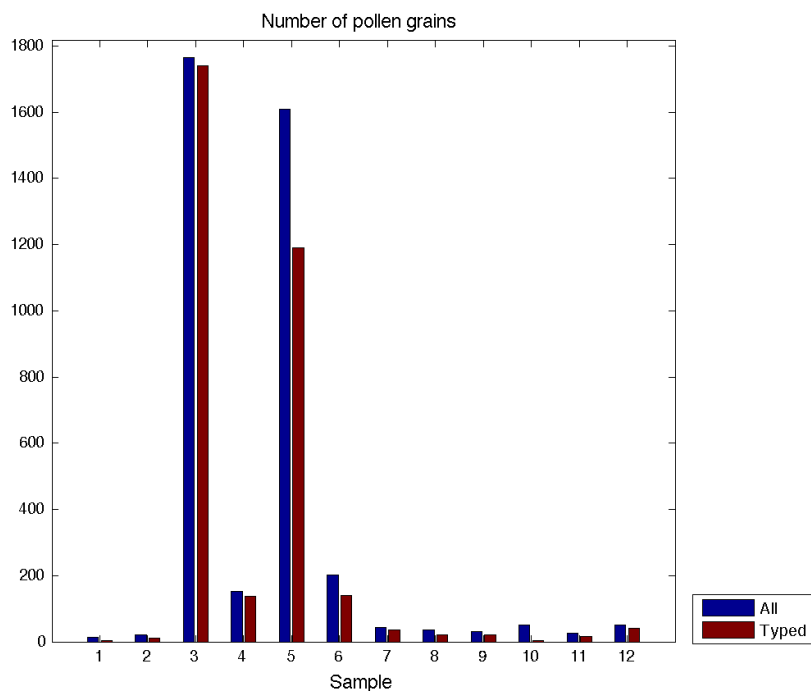


Figura 3.3: Número de granos detectados e identificados de forma manual

### 3.3. Análisis de distribución polínica anotada manualmente

El análisis de las características de los granos detectados manualmente por los expertos se ha realizado solo para aquellos granos que han sido identificados. De esta forma, se evitan posibles distorsiones por si se introduce por error un punto que no es grano de polen. En la figura 3.3 se muestra la cantidad de granos detectados en azul y de identificados en rojo. La diferencia entre ellos, en algunas muestras, es significativa y se puede deber a dos razones, a que la imagen no es nítida y por tanto no se diferencian las características del grano para su identificación, o a duda entre dos tipos, ya que los expertos lo marcan como tipo 0 para identificarlo posteriormente. Los muestras 1, 5 y 10 presentan mayor diferencia entre el número de granos detectados y el número de granos identificados porque no han sido analizadas a 400 aumentos, por lo que no ha podido mejorarse su identificación.

De los 38 tipos polínicos previstos (ver tabla A.1), se han identificado 24 tipos en las muestras analizadas (ver tabla D.1). En la gráfica 3.4 se muestra la distribución de cada tipo de grano por muestra. El tamaño de los círculos indica la cantidad de granos identificados. Se observa que aparecen distintos tipos en cada muestra. Para ver más claramente la cantidad de granos de polen totales por tipo, se presenta en

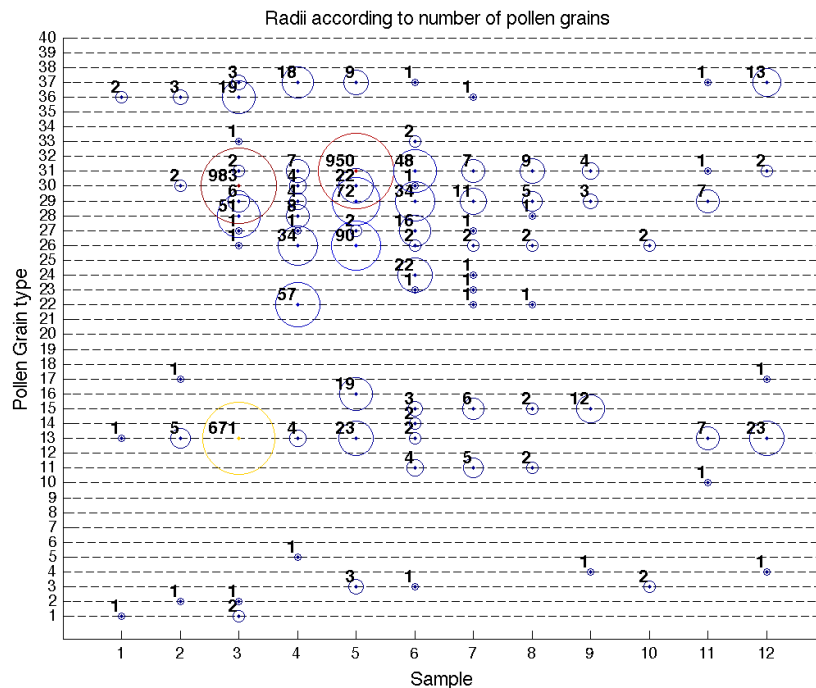


Figura 3.4: Distribución anual de tipos polínicos en el año 2012, con radio en función de la cantidad de granos identificados

la figura 3.5 una gráfica de barras donde se aprecia que hay tipos que aparecen solo una única vez y otros hasta más de 900 veces en una misma muestra. Los tipos que más se identifican en las muestras, de forma general, son los tipos polínicos 13, 30 y 31 que corresponden a Cupresáceas, Populus y Quercus respectivamente.

El estudio de la presencia de los tipos polínicos a lo largo del año (cada muestra corresponde a un mes del año) se muestra en las figuras 3.6 y 3.7, en ellas se observa que se produce un incremento de la variabilidad polínica en la primavera con un máximo en el mes de Junio. Los tipos que más aparecen a lo largo del año son 13, 26, 29 y 31, que corresponden a los tipos polínicos Cupressaceae, Pinus, Poaceae y Quercus. La razón está en que son tipos polínicos que presentan un periodo de polinización largo, de varios meses, y por tanto, aparecen en el aire en diferentes muestras.

### 3.4. Análisis de características visuales

El análisis de características de color, forma y textura se realiza para los granos de polen que aparecen al menos en 10 ocasiones, de forma que se pueda tener un conjunto de datos representativo de la variabilidad polínica, ya que un mismo



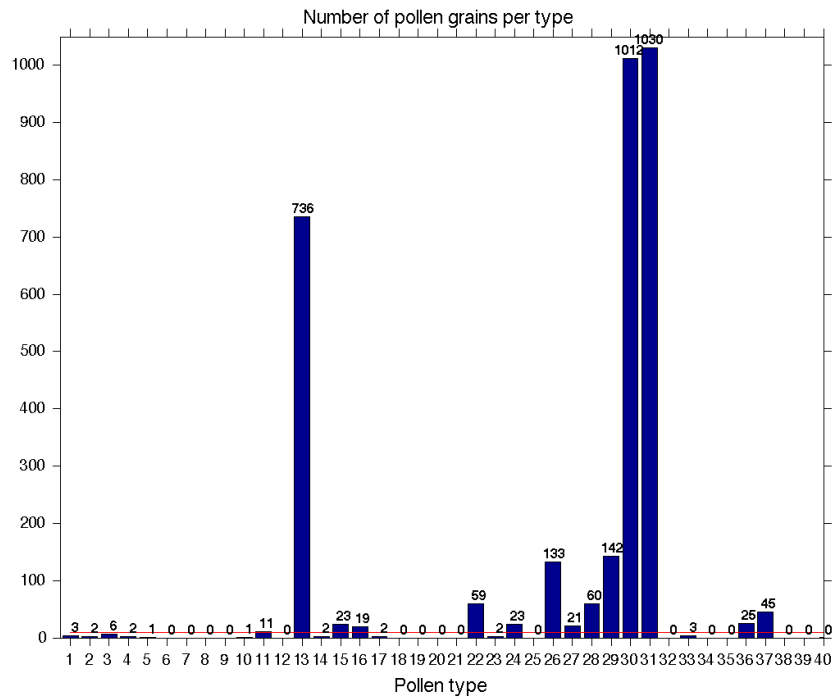


Figura 3.5: Número de granos identificados de forma manual por tipo

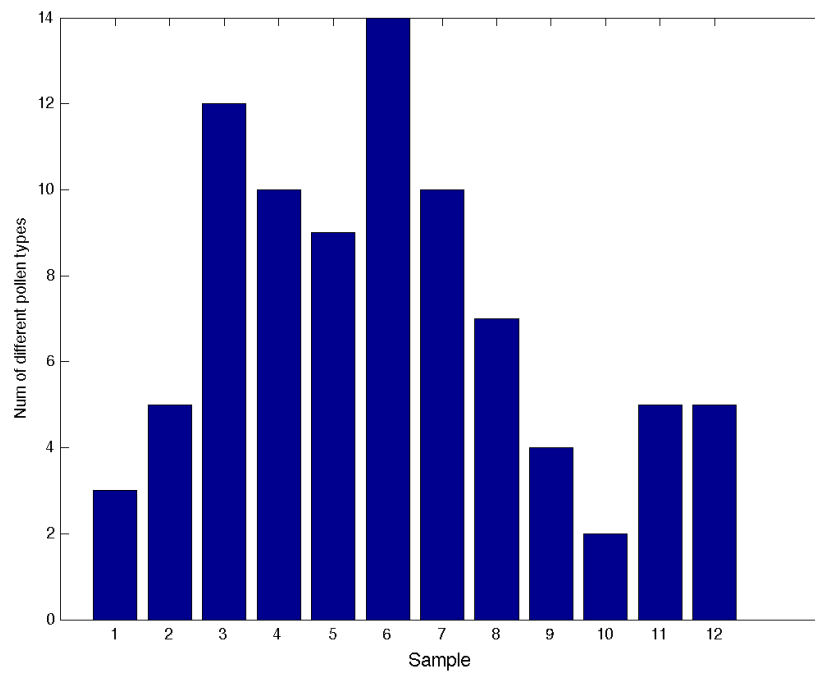


Figura 3.6: Distribución de tipos polínicos a lo largo del año 2012 por muestra

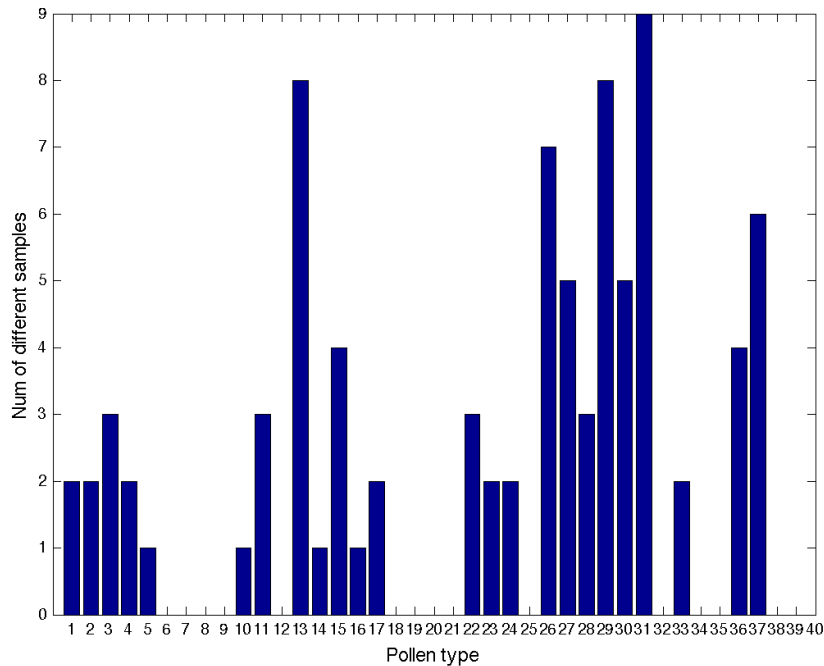


Figura 3.7: Número de meses de polinización por tipo de grano de polen detectado a lo largo del año 2012

tipo de grano varía tanto en su forma como en su color en función de su plano de digitalización o su grado de madurez. En la figura 3.5 se observa que en las muestras analizadas solo hay 13 tipos de 38 posibles, que aparecen un mínimo de 10 veces, que es lo que marca la línea roja horizontal.

### 3.4.1. Color

La delineación a partir del punto semilla anotado por el experto obtiene un conjunto de valores de pixel que se analizan en dos espacios de color. En el espacio RGB, que es el formato original de la imagen, y en el espacio HSV, que es una transformación no lineal del espacio de color RGB donde el color rojo se encuentra en valores bajos y altos del canal H. En la figura 3.8 se muestra la media (símbolo 'o') y la desviación típica (línea vertical) de cada componente de color, en el espacio RGB, por cada tipo detectado que aparece en al menos 10 ocasiones. Como era de esperar el canal rojo presenta unos valores más elevados, debidos a la tinción con fucsina, pero no se observa un patrón claro para la automatización de su detección. Por contra, en la figura 3.9 donde se muestra la media y la desviación típica en el espacio HSV, sí que se observa una clara diferencia entre las diferentes componentes. El valor medio de la segmentación de grano de polen en el canal H (matiz) y en el

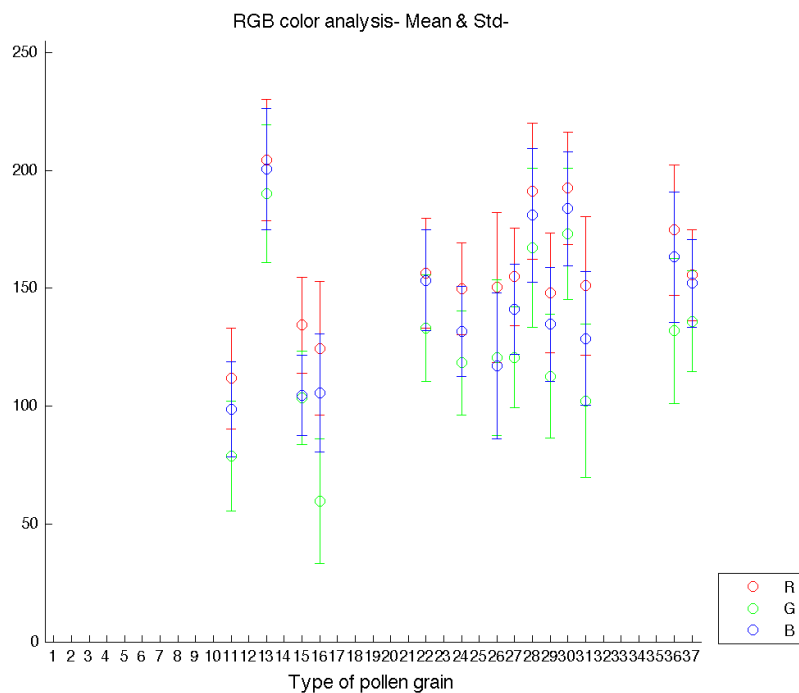


Figura 3.8: Análisis de color en espacio RGB de los tipos polínicos que aparecen en al menos 10 ocasiones

canal S (saturación) no se solapan, H tiene valores altos y S bajos. El valor medio del canal V (valor) siempre es mayor de 0.4. A pesar de que la variabilidad de los niveles del canal H es alta (debido a que los granos de polen tienen un rango amplio de color rosáceo, desde intenso hasta muy claro), se espera que el análisis en el espacio HSV obtenga buenos resultados en la detección automática de granos de polen basado en color.

También se realiza un análisis de la variabilidad de color de los granos de polen inter-muestras en el espacio de color HSV (ver figura 3.10). Se analizan los tipos polínicos que aparecen en al menos seis muestras, que son los tipos 13, 26, 29 y 31. Se observa que el color no se mantiene constante en las diferentes muestras. Existe una mayor similitud entre los granos que aparecen en muestras pertenecientes a los meses de primavera frente a los valores de los meses de otoño. Este hecho se debe a que la coloración con fucsina de las exina depende de la madurez del grano de polen. También se debe a que la digitalización de la muestra depende de la luz de la lámpara del microscopio y de la cantidad de partículas presentes en la muestra, generando una imagen con diferente tonalidad.

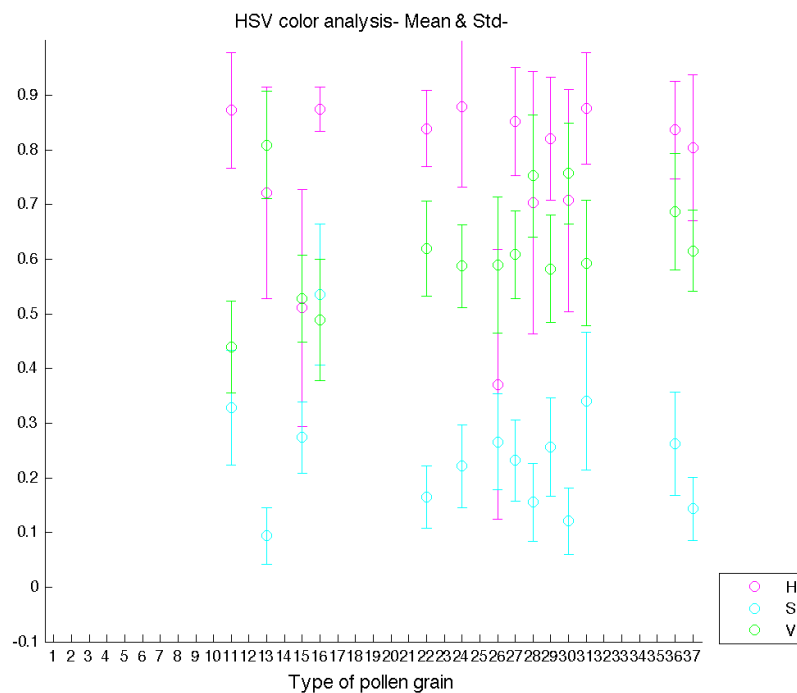


Figura 3.9: Análisis de color en espacio HSV de los tipos polínicos que aparecen en al menos 10 ocasiones

### 3.4.2. Textura

La textura mide la relación entre los diferentes elementos de una imagen y depende del tamaño la vecindad elegida. El análisis estadístico de textura se realiza mediante la entropía, que es una medida de primer orden, y la matriz de co-ocurrencia del nivel de gris (GLCM) (Haralick et al., 1973), que es una medida estadística de segundo orden. La entropía mide la uniformidad del histograma mientras que la GLCM mide la similitud entre píxeles vecinos. La GLCM es una matriz de frecuencias con la que un píxel con un nivel de gris (i) aparece en una relación de espacio específica con otro píxel de nivel de gris (j) en función de la vecindad elegida y el ángulo (ver figura 3.11).

En este trabajo se analiza la textura en tres anillos concéntricos, de tamaño variable en función de la delineación del grano de polen (ver figura 3.12). En el primer anillo se analiza el interior de grano, en el segundo la frontera y en el tercer anillo el exterior.

#### Entropía

La entropía informa sobre la distribución de niveles de gris de la imagen. A mayor entropía mayor número de niveles de gris.

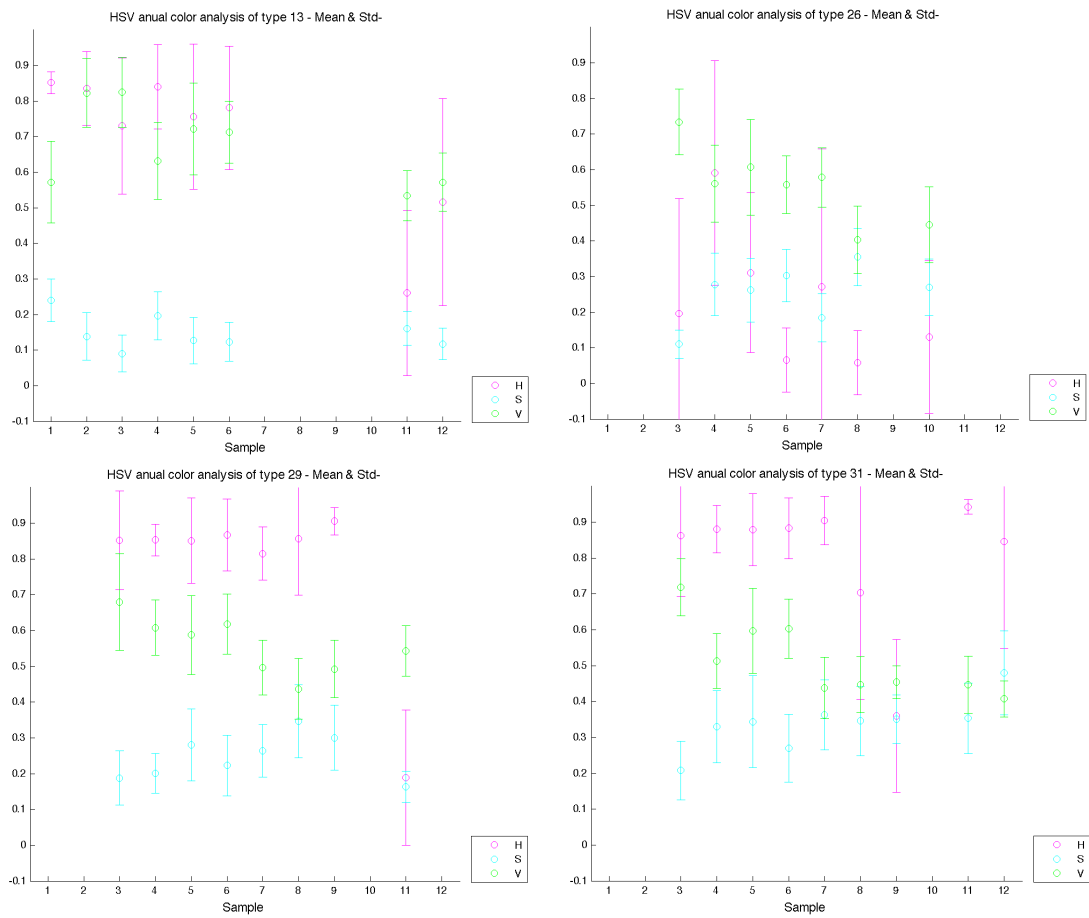


Figura 3.10: Análisis de color inter-muestra de los granos de polen en HSV

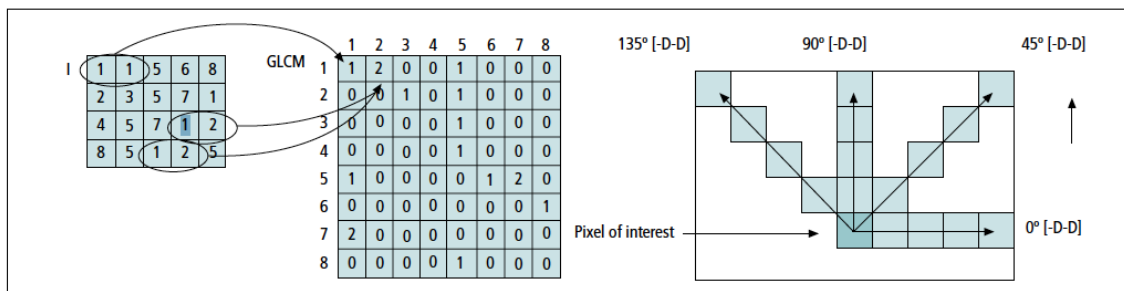


Figura 3.11: Proceso de generación de matriz GLCM con vecindad 1 y ángulo 0°

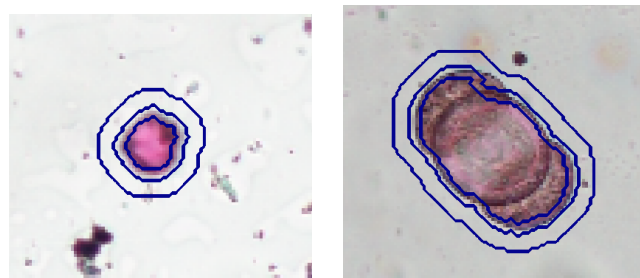


Figura 3.12: Disposición de los anillos concéntricos para análisis de textura

$$ENTROPIA = \sum_i p(i) * \log_2(p(i))$$

En la figura 3.13 se muestra mediante un diagrama de cajas<sup>1</sup> la variación de la entropía en cada anillo por cada tipo de grano de polen que aparece en al menos 10 ocasiones, . Se observa que, en general, en el interior del grano la variación de la textura es inferior que en el resto de anillos. En el anillo intermedio la entropía es más elevada ya que se analiza el borde de los granos de polen, y es una zona que puede contener pixeles del interior, la corteza y el exterior del grano. El anillo exterior analiza el contenido de la imagen que existe alrededor del grano. Valores bajos de entropía indican que la muestra es limpia, ya que todos los pixeles pertenecen al fondo, mientras que niveles altos, indican que existen múltiples objetos cercanos al grano. En general, los granos se encuentran en imágenes con bastantes partículas salvo en los tipos 22 y 37 (Moraceae y Urticaceae, respectivamente). Los granos de polen de estos tipos pertenecen en su mayoría a las muestra 4 y 12 que son muestras muy limpias (ver figura 5.5). El caso 37 (Urticaceae) muestra una entropía más baja en todos los anillos debido a que es un tipo con diámetro pequeño y presenta una pigmentación muy clara, por lo que no existe mucha variabilidad con el fondo. El caso 26, es el grano tipo Pinus, presenta una mayor entropía en el anillo interno e intermedio debido a su morfología que presenta dos sacos aéreos. Los anillos puede contener partes de la intina, la exina y de los sacos.

### Matriz de co-ocurrencia del nivel de gris

Se calcula la matriz de co-ocurrencia de niveles de grises con un tamaño de ventana de 8 x 8 y una vecindad [0 1], es decir, se escala la imagen a 8 niveles de grises y se analiza el pixel vecino de la derecha ( 0º) que tiene distancia 1. A partir de la GLCM se pueden calcular características de textura como la homogeneidad local, el contraste, la energía o la correlación. Entre estas características solo se describirán, en este documento, el contraste y la correlación.

- Contraste:

$$CONTRASTE = \sum_{i,j} |i - j|^2 p(i, j) \quad Rango = [0 (Ng - 1)^2]$$

<sup>1</sup>El diagrama de cajas es un gráfico que muestra información sobre los valores mínimo y máximo, los cuartiles, la mediana y la media. También informa sobre la existencia de valores atípicos y la simetría de la distribución. El símbolo 'o' indica la media, 'v' es la mediana y la caja los cuartiles 25 (q1) y 75 (q3). En línea discontinua se muestra el máximo y el mínimo como 1.5 veces del rango intercuartil (q3-q1). Los valores atípicos se muestran con puntos.

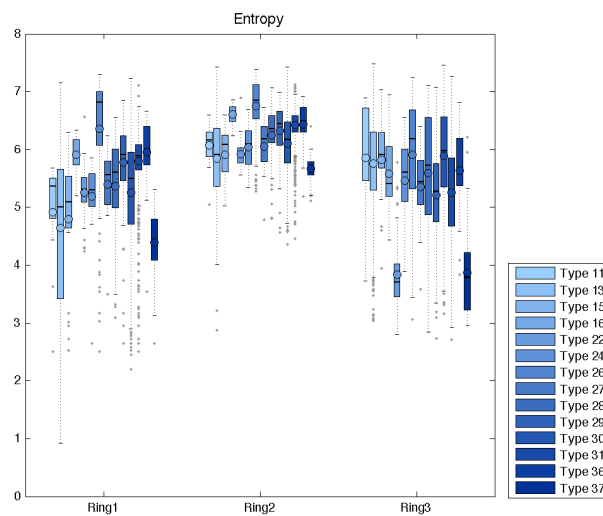


Figura 3.13: Análisis de textura en anillos concéntricos: Entropía

con  $Ng = NumNivelesGris$

El contraste mide la variación local en una imagen. A mayor cambio entre píxeles de la imagen mayor es su contraste. En la figura 3.14 se muestra la variación del contraste en cada anillo por cada tipo de grano de polen que aparece en al menos 10 ocasiones. Se observa que no hay muchos cambios en la imagen en el análisis de los anillos interno e intermedio y que aparecen valores más elevados en el anillo exterior, debido a la aparición de otros objetos cercanos al grano. Un caso especial es el grano tipo Pinus que presenta valores más elevados que es resto. Debido a su morfología la delineación a partir del punto semilla no se realiza correctamente (ver figura 3.15), por ello hay una gran variabilidad de valores en los píxeles de los anillos, sobre todo en el anillo exterior.

- Correlación:

$$CORRELACION = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad Rango = [-1 \ 1]$$

La correlación de la GLCM mide la similitud entre píxeles vecinos. Si los píxeles son idénticos toma valor 1 mientras que si todos los píxeles vecinos son distintos toma valor -1. En la figura 3.16 se muestra la variación de la correlación en cada anillo por cada tipo de grano de polen que aparece en al menos 10 ocasiones. Se observa que, en general, existe una alta similitud entre píxeles vecinos. La variación entre píxeles es suave. La correlación es menor en el interior y el borde del grano de polen, ya que su estructura física es porosa y varía en función del nivel de enfoque.

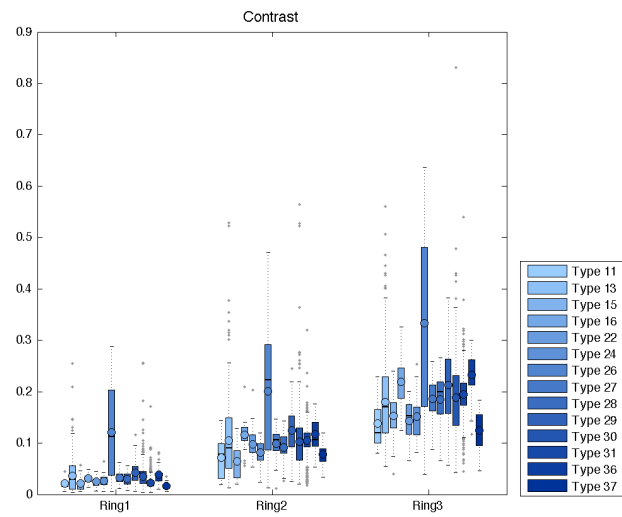


Figura 3.14: Análisis de textura en anillos concéntricos: Contraste



Figura 3.15: Anillos defectuosos en el grano de polen tipo Pinus



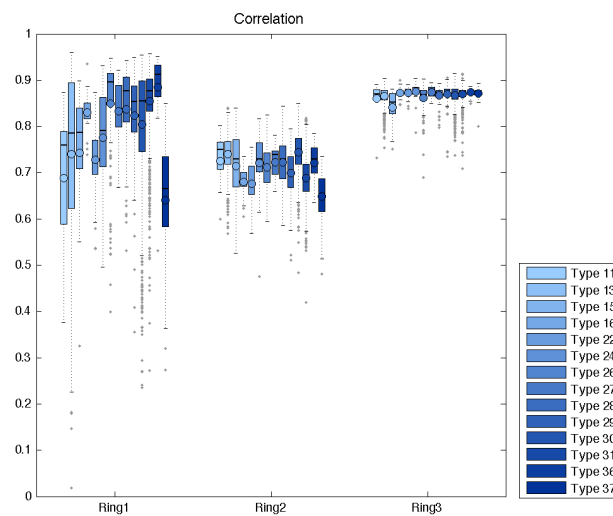


Figura 3.16: Análisis de textura en anillos concéntricos: Correlación

En el exterior la correlación es alta ya que la mayoría de los píxeles son similares y pertenecen al fondo.

### 3.4.3. Forma y tamaño

Según la bibliografía, los granos de polen son, en su mayoría, objetos esféricos de tamaño variable entre 10 y 250  $\mu\text{m}$ . A partir de la delineación generada de forma automática a partir del punto semilla anotado por el experto, se analiza experimentalmente el tamaño y forma de los granos de polen que aparecen en las muestras analizadas en al menos 10 ocasiones. En la figura 3.17 se muestra la variación del área por cada tipo polínico estudiado. Se observa que los granos de tipo 26 (*Pinus*) presentan un área muy elevada, son granos de tamaño grande de unos 50-100  $\mu\text{m}$ , mientras que los de tipo 37 (*Urticaceae*) tienen un área muy pequeña, de unos 10  $\mu\text{m}$ . El resto son granos de tamaño medio, entre 10 y 50  $\mu\text{m}$ .

El análisis de la longitud del diámetro del eje mayor y del eje menor de la delineación de cada grano de polen se muestra en la figura 3.18. Se observa que la medida del eje mayor y del eje menor son similares, ya que los objetos analizados son, en general, circulares.

En las gráficas anteriores, figuras 3.17 y 3.18, se observan valores atípicos, fundamentalmente en el tipo 31. Se debe a errores en la delineación automática a partir de los puntos semilla. En general, se obtienen buenos resultados con la mayoría de los puntos semilla, pero hay algunos casos en los que la segmentación no encuentra ningún objeto, por lo que se asigna una delineación de un círculo 8  $\mu\text{m}$

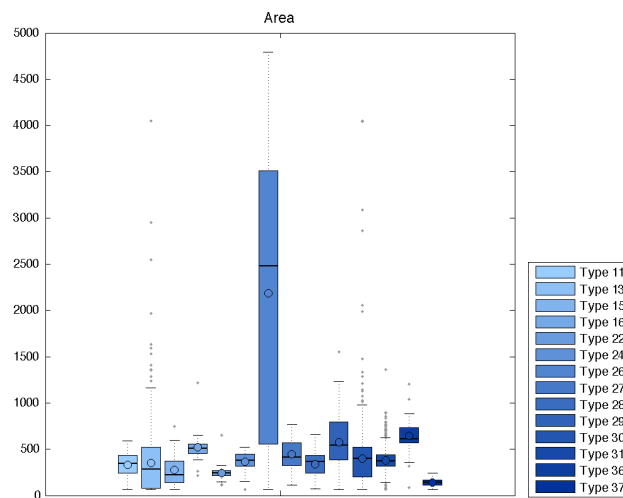


Figura 3.17: Tamaño de los granos de polen por tipo: Area

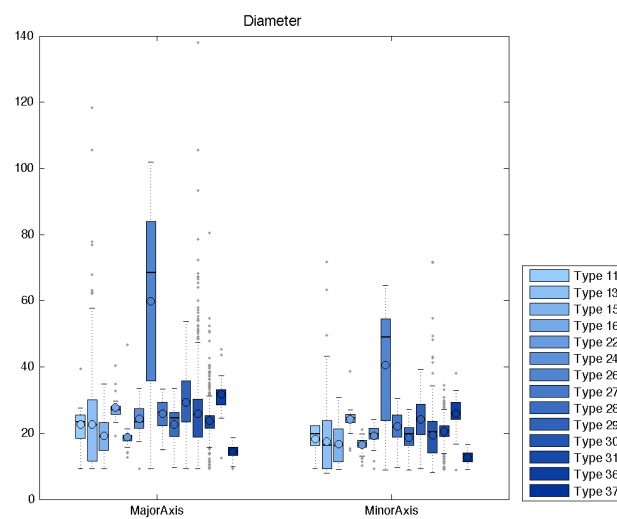


Figura 3.18: Tamaño de los granos de polen por tipo: Diámetro eje mayor y menor



Figura 3.19: Outliers a partir de análisis del tamaño

de diámetro centrado en el punto semilla. En otros, se añaden estructuras rosáceas que se encuentran anexas al grano de polen analizado aumentando su tamaño (ver figura 3.19). La influencia de este tipo de errores se minimizan al analizar múltiples granos de polen de un mismo tipo, por ello, en este trabajo se han estudiado los granos que aparecen al menos en 10 ocasiones.

### 3.5. Resumen del capítulo

Se han analizado 12 muestras, cada una procedente de un mes del año 2012 y procedentes de un captador situado en Toledo, con una digitalización con enfoque automático y de forma manual por diferentes expertos palinólogos. A partir de este análisis, se ha mejorado el conocimiento de las características visuales de los granos de polen (color, forma tamaño, textura, etc.) permitiendo desarrollar un sistema automático para la detección e identificación de granos de polen. Se ha comprobado que existe una gran variabilidad de las características de color en los diferentes tipos polínicos identificados, incluso dentro de un mismo tipo polínico. El espacio de color que más información aporta es el espacio HSV que gracias a la transformación no lineal que utiliza divide la imagen en distintos colores. En la detección de granos de polen se buscan objetos de color rojizo que se encuentran en niveles bajos y altos del canal H.

En las imágenes diarias a baja resolución los expertos han detectado 4061 posibles granos de polen en las 12 muestras analizadas. Por cada posible grano no identificado, este se digitaliza a una mayor resolución y a diferentes planos de enfoque. Se

descartan puntos que no son granos de polen y, al final, quedan 3999 granos de polen con 24 de 38 tipos posibles, de los cuales 636 quedan sin identificar. En el anexo D se adjunta una imagen de cada tipo polínico identificado junto a sus características físicas más destacadas.

# Capítulo 4

## Sistema inteligente remoto de digitalización, detección e identificación de granos de polen

### 4.1. Introducción

El protocolo actual para el recuento de granos de polen aerotrasportados se realiza a partir de fragmentos diarios de una cinta adhesiva que tiene adheridos granos de polen mediante la utilización de captadores de aire por succión. Cada fragmento se prepara y se analiza de forma manual. Este es un proceso lento y costoso pues requiere de profesionales entrenados para detectar e identificar granos de polen en muestras que contienen múltiples partículas.

Para ayudar en la tarea de detección e identificación, se propone un nuevo método que automatice el proceso, de forma que los expertos solo tengan que identificar aquellos casos en los que el sistema automático no encuentre solución.

Con la metodología actual, la detección de granos, por parte de los expertos, no se realiza de toda la muestra completa, ya que su extensión es muy elevada. Se analiza, al menos, un 10 % de la muestra siguiendo las directrices de la REA y se aproxima el contenido polínico a toda el fragmento analizado. Con el sistema automático propuesto se puede analizar el 100 % de la muestra. Para ello, es necesario digitalizar la muestra de forma completa para posteriormente ser analizada por un sistema de detección e identificación de granos de polen.

En la figura 4.1 se muestra el esquema de funcionamiento del sistema desarrollado.

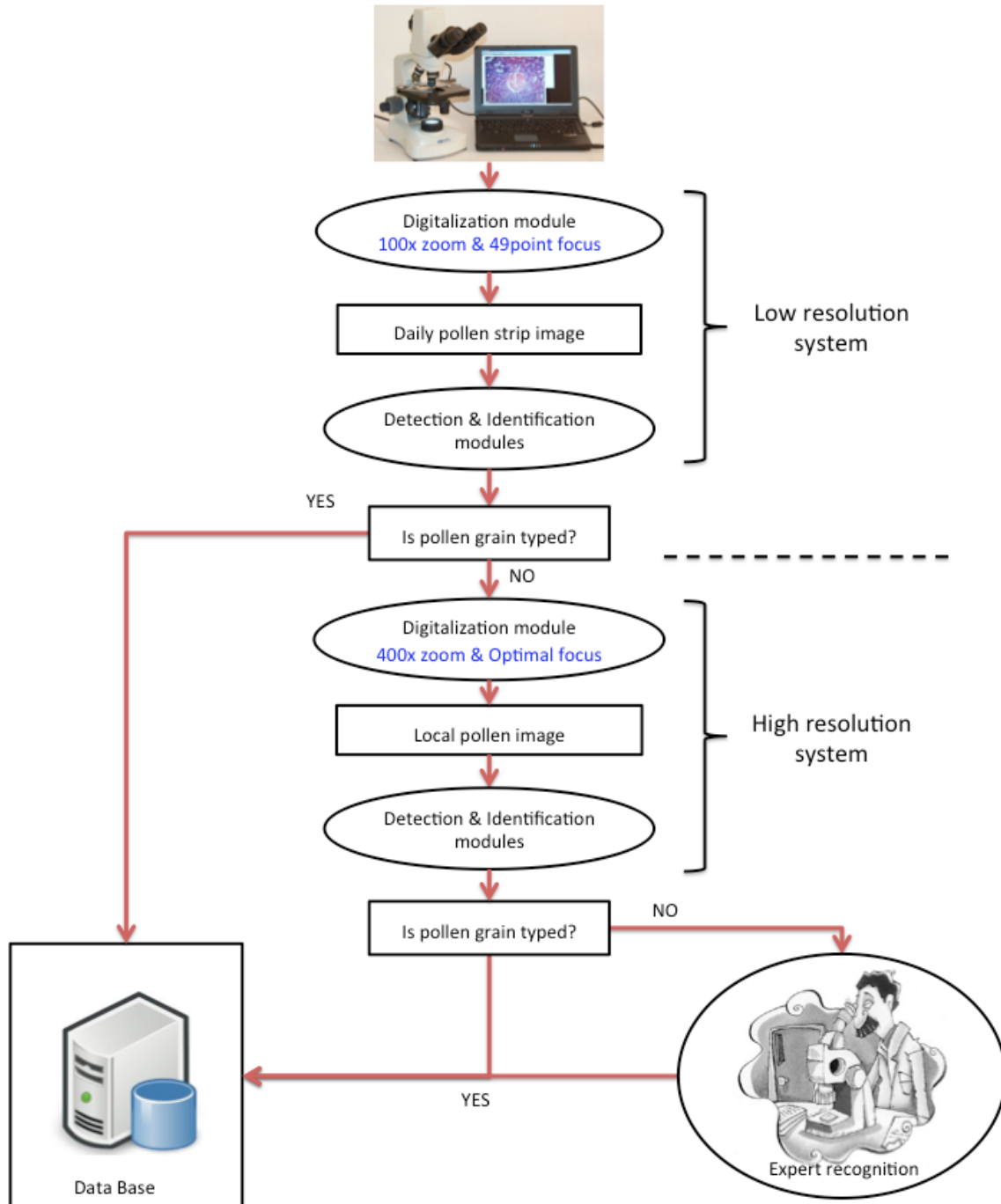


Figura 4.1: Esquema de funcionamiento del sistema semiautomático de análisis

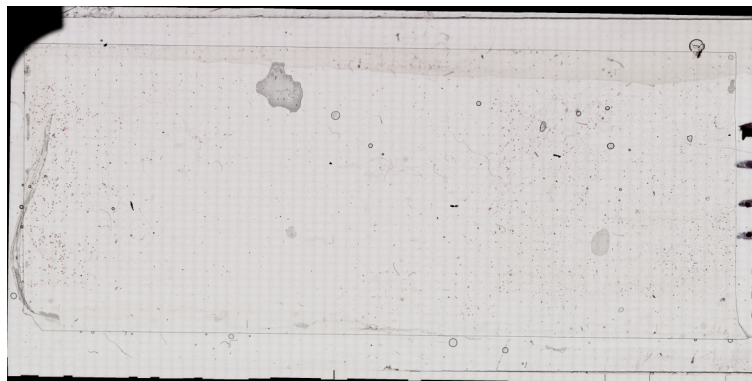


Figura 4.2: Muestra diaria digitalizada a baja resolución

## 4.2. Sistema a baja resolución

### 4.2.1. Módulo de digitalización

La adquisición de toda la muestra, cuyo tamaño es de 48 x 19 mm, se realiza en una primera etapa a baja resolución, con un aumento de 100 magnitudes mediante un mosaico de subimágenes de 640 x 480 píxeles en RGB con un tamaño de 1.34  $\mu\text{m}/\text{pixel}$ . El tamaño resultante de una muestra diaria digitalizada es de 1GB aproximadamente, unos 36480 x 14400 píxeles (ver figura 4.2).

La imagen obtenida tras la digitalización de la muestra contiene el fragmento de 24 horas de captación y las marcas añadidas a la muestra para el recuento mediante barridos horizontales según indicaciones de la REA. En la figura 4.3 se presenta la zona digitalizada de la muestra con un rectángulo rojo y las marcas de las cuatro líneas horizontales. Para utilizar una imagen que se ajuste al fragmento de la cinta adhesiva, se seleccionan los bordes de la cinta y la situación de las líneas de barrido. Se guardan las posiciones de las líneas para comparar el método tradicional con el método digital, tanto de forma manual como automática.

En los resultados experimentales se ha observado que existe una variación del enfoque óptimo en la muestra debida a varios motivos:

1. Las partículas en suspensión que se adhieren a la cinta transparente tienen distintas formas y tamaños, lo que impide obtener una buena digitalización de toda la preparación con una única posición de enfoque (ver fig. 4.4).
2. Existe una pequeña inclinación de la base motorizada donde se sitúa el portamuestras respecto a la cámara (fig. 4.5).

Para solucionar este problema se analiza una malla de 49 puntos equiespaciados (ver fig. 4.6), que permiten obtener el plano de menor error medio de enfoque. Con

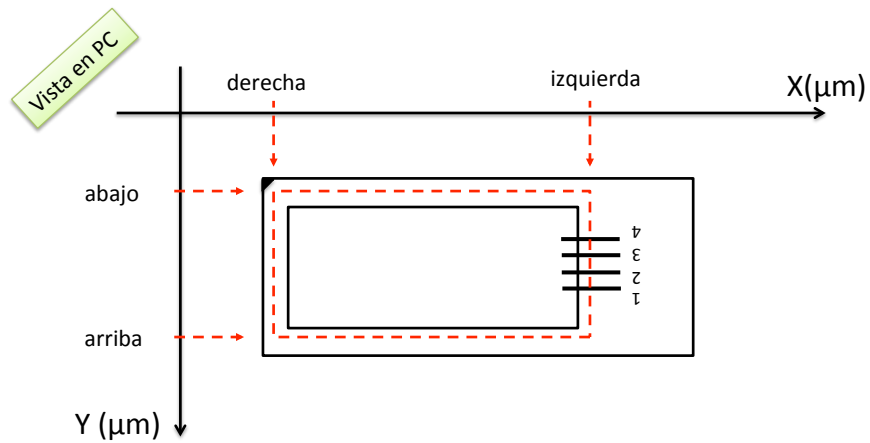


Figura 4.3: Situación de la muestra digitalizada sobre la preparación respecto al origen de coordenadas de la base motorizada

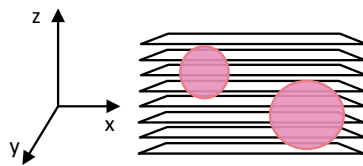


Figura 4.4: Situación de los granos en el eje z de la cinta adhesiva

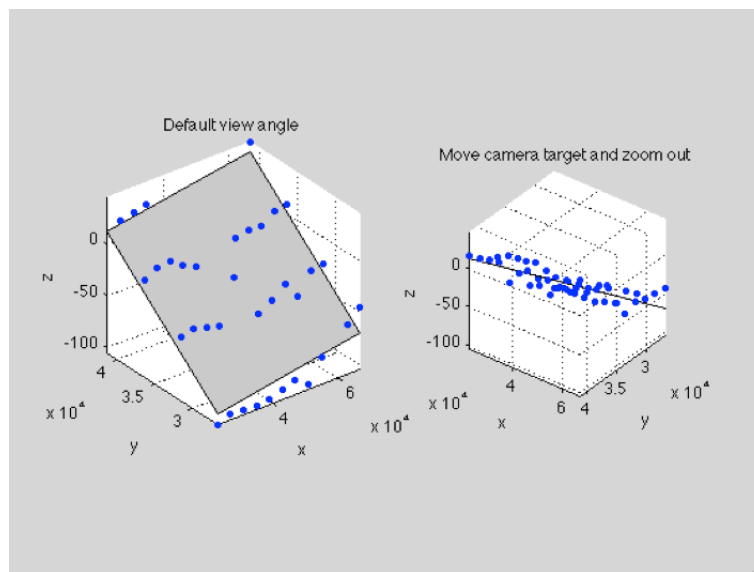


Figura 4.5: Plano de enfoque que mejor se adapta a la muestra



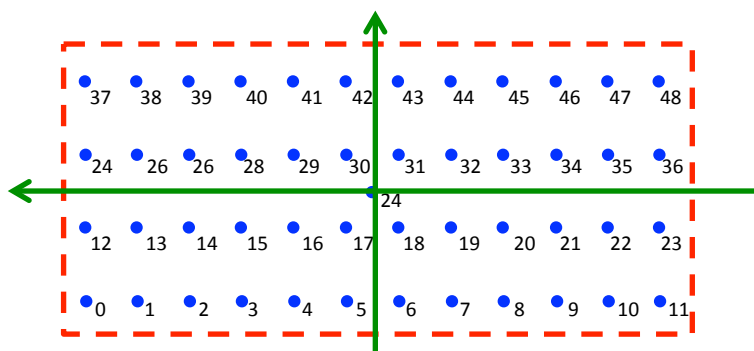


Figura 4.6: Malla de 49 puntos equiespaciados para analizar enfoque

este plano se ajusta el eje Z durante la digitalización total de la preparación, proceso que dura aproximadamente unos 60 minutos.

Debido a la configuración del hardware y software utilizado, la imagen digitalizada de la muestra de polen se almacena en memoria rotada  $180^\circ$  respecto a lo que se observa al microscopio (ver figura 4.3). Por tanto, hay que tenerlo en cuenta a la hora de dividir verticalmente la muestra en función de la hora, ya que serían intervalos decrecientes de izquierda a derecha.

#### 4.2.2. Modulo de detección

La primera fase de análisis consiste en la búsqueda de objetos con características similares a los granos de polen en toda la muestra digitalizada a baja resolución (100 aumentos). Como se comentó en el capítulo anterior, los granos de polen tienen una forma circular, pueden contener aperturas y su tamaño varía entre 10 y 250  $\mu\text{m}$ . Además, presentan un color rosáceo debido a la tinción con fucsina que se realiza en la preparación de la cinta. La muestra contiene múltiples partículas, tanto bióticas como abióticas, que han de ser diferenciadas de los granos de polen.

Para llevar a cabo la tarea del recuento polínico sobre una muestra digitalizada, se ha desarrollado un módulo de detección de granos de polen que consta de las siguientes fases:

1. Eliminación de fondo,
2. Umbralización por color,
3. Filtrado por tamaño,
4. Análisis de forma, y
5. Filtrado mediante características de color, forma y textura de cada posible grano.

El esquema de funcionamiento del sistema de detección de objetos rosáceos con forma circular se muestra en la figura 4.7. Cada una de las etapas del módulo desarrollado se comenta a continuación con mayor detalle.

**Eliminación de fondo** La muestra digitalizada está compuesta por un fondo claro que ha de ser eliminado para llevar a cabo el proceso de detección de granos de polen. Para ello, se transforma la imagen a niveles de gris y se umbraliza en función del histograma de la imagen. Debido al gran tamaño de la imagen, la distribución del histograma de la imagen se puede aproximar a una gaussiana, con la que se calcula el umbral como la media menos una desviación típica para eliminar los píxeles de color blanco. No se utiliza un umbral fijo ya que cada muestra puede tener diferente nivel de gris en la imagen de fondo, en función de la cantidad de partículas o la iluminación de la lámpara del microscopio en la digitalización.

Además de eliminar el fondo, se eliminan aquellos objetos muy oscuros (ver figura 4.8) como pueden ser burbujas o sustancias minerales mediante un umbral fijo de valor bajo (igual a 0.2 partiendo de una imagen de niveles de gris entre 0 y 1). Este valor no puede ser alto ya que hay los granos de polen intensos que son oscuros en la imagen de grises que no se deben de perder en etapas tan tempranas.

**Umbralización por color** El rango de valores de color de los granos de polen varía desde tonos rosáceos intensos a muy pálidos, como se observa en la figura 4.9. Tras el análisis de los granos de polen realizado en el capítulo 3, se utiliza el espacio de color HSV en la búsqueda de objetos rosáceos compactos. Se parte de una imagen en el espacio HSV donde se eliminan los píxeles pertenecientes al fondo y a objetos muy oscuros. Para obtener el mayor número de píxeles rosáceos pertenecientes a un grano, se realiza un barrido del histograma en el canal H, con una ventana de anchura fija entre los rangos  $H_{min}$  y  $H_{max}$  en busca de objetos compactos con un área mayor a 5 píxeles. Se ha comprobado que los valores del canal H en los granos de polen varía desde valores bajos (granos de polen terrosos) a valores altos (granos de polen rojizos). Por ello, se concatenan los objetos compactos de las N ventanas de valor mayor y menor de H.

Tras observar los valores obtenidos en la caracterización de los granos a partir de los puntos anotados por los expertos (ver figura 3.9), el canal S se umbraliza con valores fijos entre 0.05 y 0.6 sobre la imagen normalizada entre 0 y 1.

A pesar de que no se obtuvo un patrón claro en el análisis de los granos de polen anotados de forma manual, sí que se puede obtener una relación entre los canales en RGB, es decir, se pueden buscar píxeles que tengan un valor de en el canal rojo un 15% mayor que en el canal verde y el canal azul y que el canal azul sea un 10%

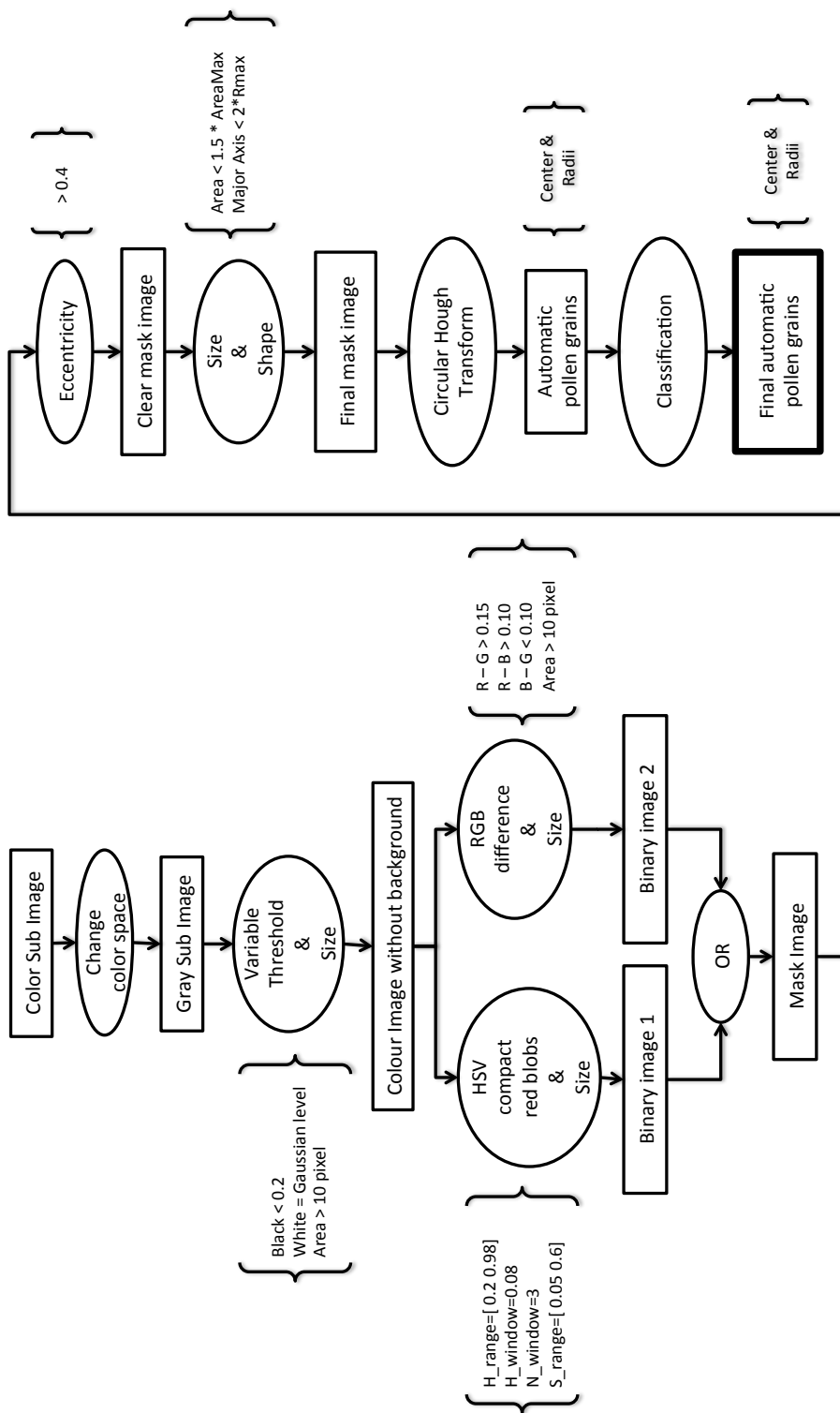


Figura 4.7: Esquema de funcionamiento del sistema de detección de granos de polen

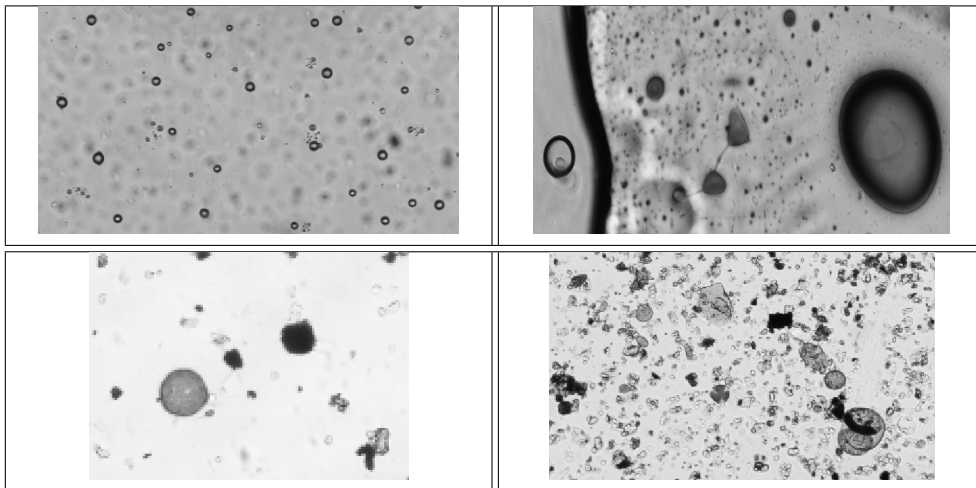


Figura 4.8: Ejemplo de objetos muy oscuros presentes en las muestras

mayor que el canal verde. Por ello, se añade al módulo de detección de granos de polen el análisis de la imagen en el espacio de color RGB tras eliminar los píxeles del fondo y objetos oscuros con umbralización fijas entre los valores de diferencias de los distintos canales.

Al final, se obtienen imágenes binarias del análisis de color en HSV y en RGB que se interseccionan y se aplican operaciones morfológicas de cierre y rellenado de agujeros para suavizar la segmentación resultante.

**Filtrado por tamaño** La segmentación de granos de polen en una imagen que contiene múltiples partículas, algunas de las cuales también presentan una coloración rosada, es una tarea compleja. En ocasiones se localiza un grano cerca de objetos vegetales (también rosáceos) que producen una segmentación con ambos objetos y de mayor tamaño. Para evitar la pérdida de estos granos (falsos negativos) se ha utilizado como área máxima el valor de un círculo de radio máximo de los tipos de polen estudiados, incrementado un 50%, es decir,  $1,5 * 2 * \pi * R_{max}$ , con  $R_{max} = 200\mu m$ . Lo mismo sucede en el análisis del eje mayor del objeto, para evitar la pérdida de granos solapados o cercanos se utiliza una restricción baja, en concreto, solo se eliminan aquellos objetos con una longitud mayor que 2 veces  $R_{max}$ . El área mínima utilizada para los objetos segmentados es de 1/4 del círculo de radio mínimo de polen, es decir,  $0,25 * 2 * \pi * R_{min}$ , con  $R_{min} = 10\mu m$ , de forma que se pueda detectar al menos una parte en los granos de polen pequeños.

**Análisis de forma** Los granos de polen son objetos cuasi-circulares con un color rosáceo muy variado. La segmentación por umbralización basada en color produce, en algunas ocasiones, la detección de solo una parte del grano, por lo que no se obtiene

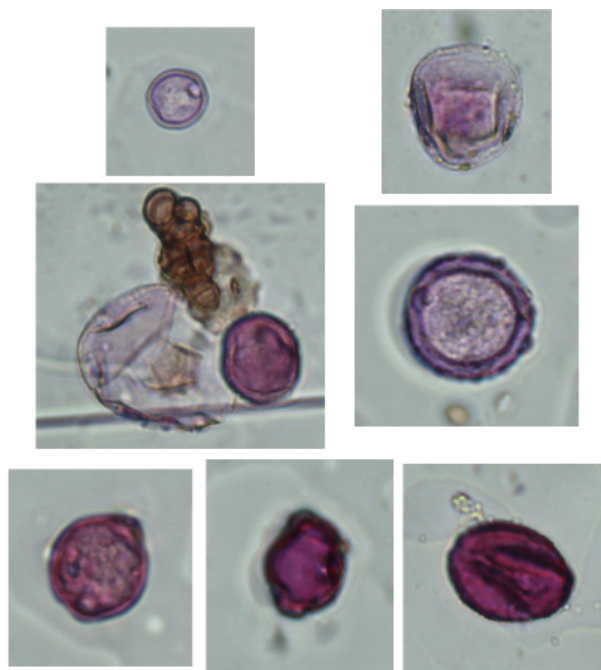


Figura 4.9: Diferente pigmentación rosácea de los granos de polen

un objeto circular compacto. Para evitar la pérdida de este tipo de granos, se utiliza un coeficiente bajo en el análisis de excentricidad de valor 0.4 donde un objeto circular perfecto tendría un valor igual a la unidad.

También se analiza la segmentación obtenida en busca de objetos circulares mediante la transformada de Hough. La transformada de Hough es un método para la detección de contornos que utiliza conocimiento de la ecuación paramétrica del objeto que se busca en la imagen. No se ve afectado por las lagunas existentes en la frontera del objeto ni por el ruido. La versión circular utiliza un acumulador de tres parámetros  $a$ ,  $b$  y  $r$ , ya que la ecuación paramétrica del círculo viene dado por  $(x - a)^2 - (y - b)^2 = r^2$ . En el caso de los granos de polen, se parte de un rango de posibles radios  $[Rmin Rmax]$  y se obtiene un lista de puntos con centro y radio del objeto detectado. Se utiliza como imagen de entrada la segmentación obtenida en la etapa anterior, mediante métodos basados en pixel. El análisis de la circularidad con la transformada de Hough permite eliminar aquellas segmentaciones que no presentan una forma circular. También, separa segmentaciones de objetos solapados o muy cercanos que se segmentan mediante un único objeto. Un ejemplo se muestra en la figura 4.10.

**Filtrado por características locales y regionales** La detección de objetos con características similares a los granos de polen genera un conjunto de puntos muy elevado, muchos de los cuales no son verdaderos granos de polen, es decir son falsos



Figura 4.10: Segmentación mediante Transformada de Hough de granos adyacentes

positivos. Para minimizar su número se ha entrenado un clasificador mediante 11 características de color, forma, tamaño y textura con 7998 instancias, de las cuales 3999 son de clase 1 (Si es grano). El tipo de clasificador elegido es un árbol de decisión tipo C.45 (en Weka es J48). Se utiliza este tipo de clasificador por su sencillez, a pesar de obtener peores resultados que una red neuronal. La información de las ramas del árbol permiten ajustar los parámetros del sistema de detección y explicar a los expertos palinólogos el funcionamiento del sistema automático. En el anexo B se explica en detalle las características visuales analizadas de los granos de polen que se utilizan en este clasificador. De todas las variables analizadas, se utilizan las que presentan una mayor influencia en la clase de salida. Tras probar varias configuraciones se utilizan 11 variables: el área (variable de tamaño), la excentricidad (variable de forma), el valor medio de canal H en todo el objeto detectado, en el interior y en la corteza junto al valor medio del canal V en todo el objeto y el valor medio del canal verde en el exterior de objeto como variables de color y, por último, como variables de textura se utiliza la correlación y la entropía en el interior del objeto, la homogeneidad en la frontera y de nuevo la correlación en el exterior.

### 4.2.3. Módulo de identificación

La identificación de granos de polen en la imagen a baja resolución es difícil, ya que la imagen no siempre tiene suficiente información para detectar la pared o las aperturas, características fundamentales para discriminar los tipos. Para realizar una primera identificación con imágenes a baja resolución, se ha entrenado un clasificador a partir de las características de los granos de polen identificados mediante el método digital manual. De los 38 tipos utilizados, solo se han detectado un número suficientemente elevado en los tipos 13, 26, 29, 30, 31. Por lo que este clasificador preliminar solo puede discriminar en esta fase 5 tipos polínicos. En la figura 4.11 se muestra el resumen de resultados tras evaluar el clasificador. El porcentaje de

```

Correctly Classified Instances      2551          75.8549 %
Incorrectly Classified Instances    812           24.1451 %
Kappa statistic                    0.6813
Mean absolute error                 0.067
Root mean squared error             0.2341
Relative absolute error              35.2388 %
Root relative squared error         75.9214 %
Total Number of Instances          3363

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      0.44    0.034   0.435     0.44    0.438     0.693     0
      0.751   0.061   0.775     0.751   0.763     0.871     13
      0.678   0.006   0.678     0.678   0.678     0.886     22
      0.669   0.01    0.742     0.669   0.704     0.837     26
      0.2     0.014   0.203     0.2     0.202     0.649     28
      0.331   0.024   0.379     0.331   0.353     0.65      29
      0.784   0.108   0.757     0.784   0.77      0.83      30
      0.906   0.048   0.892     0.906   0.899     0.937     31
Weighted Avg.  0.759   0.065   0.756     0.759   0.757     0.854

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  <-- classified as
84 11  8  7  3 15 23 40 | a = 0
15 553 1  3  9  9 144  2 | b = 13
 8  3 40  0  0  2  5  1 | c = 22
 7  4  1 89  0  8 13 11 | d = 26
 0  6  0  1 12  2 34  5 | e = 28
26  9  4  5  2 47 16 33 | f = 29
19 122 3  7 29 18 793 21 | g = 30
34  6  2  8  4 23 20 933 | h = 31

```

Figura 4.11: Resultados de entrenamiento y test de árbol de decisión para identificación polínica a 100 aumentos

instancias mal clasificadas es elevado por lo que es necesario mejorar el modelo.

## 4.3. Sistema a alta resolución

### 4.3.1. Módulo de digitalización

Una vez se tiene el conjunto de puntos donde se sitúa un posible objeto similar a un grano de polen que no ha sido identificado a baja resolución, se procede a la captura de imágenes tridimensionales a alta resolución. Se obtiene una pila de imágenes, con salto en función del diámetro del objeto localizado, a 400 aumentos de 640 x 480 píxeles en RGB y un tamaño de 0.34  $\mu\text{m}/\text{pixel}$  (ver figura 4.12). Se localiza la imagen mejor enfocada dentro de la pila de imágenes capturadas mediante un análisis de bordes y textura, para posteriormente ser analizada y extraer sus características visuales para su identificación.

### 4.3.2. Módulo de detección

La imagen obtenida tras la digitalización a 400 aumentos contiene al objeto rosáceo y circular en el centro de la imagen. Para obtener la delineación de grano de polen bajo análisis se realiza una segmentación de la imagen mediante el mismo sistema de detección que se se utiliza con la imagen a 100 aumentos (baja resolución).

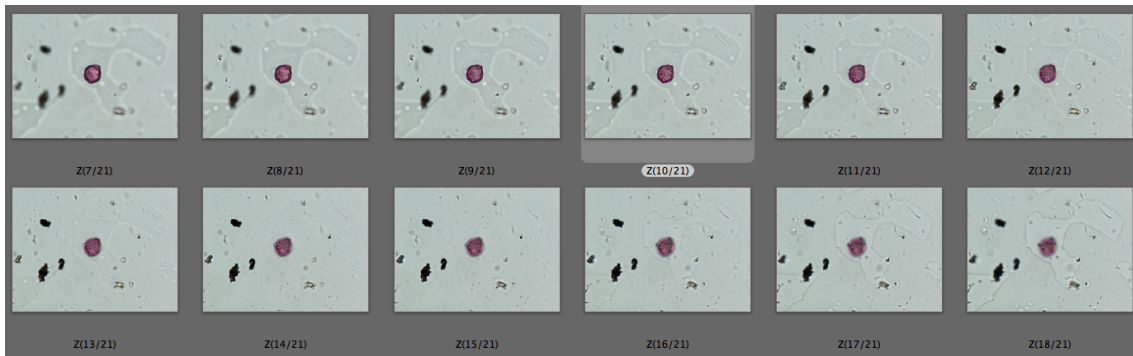


Figura 4.12: Pila de imágenes que componen la imagen 3D del grano de polen

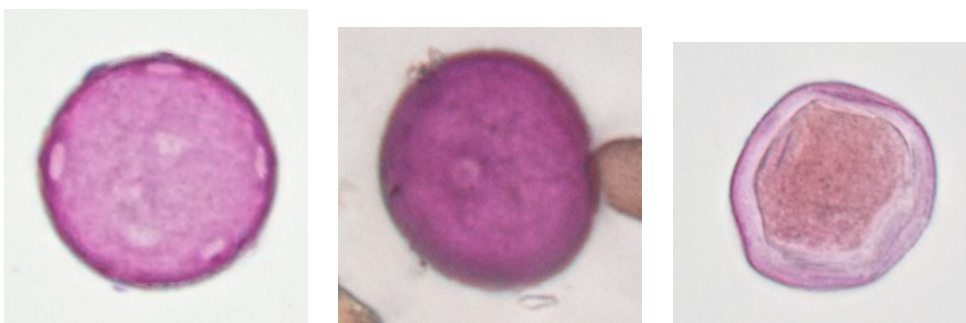


Figura 4.13: Granos de polen *Plantago*, *Poaceae* y *Populus* a 400 aumentos

Como se conoce la posición y el tamaño del objeto, se añade a la segmentación un círculo centrado en el punto semilla  $c$  y radio  $r$ , que son variables que han sido obtenidas de forma automática en la etapa anterior a 100 aumentos.

La nueva imagen del posible grano de polen tiene una mayor resolución que permite observar las diferentes estructuras que componen el grano de polen. Mediante un análisis del histograma del objeto segmentado se detectan los diferentes elementos del objeto, que se utilizan en la fase de identificación.

### 4.3.3. Módulo de identificación

La identificación de granos de polen a alta resolución se realiza mediante un clasificador desarrollado por (Prieto et al., 2011) a partir del aprendizaje de las características de color, tamaño, forma, aperturas y textura. Por ahora, el clasificador desarrollado solo identifica tres familias de granos de polen: *Plantago*, *Poaceae* y *Populus*. Según (Prieto et al., 2011) se obtiene una tasa de reconocimiento del 99.11 % mediante redes neuronales y un 98.21 % mediante un árbol de decisión J48 con muestras puras.



## 4.4. Sistema de verificación

Una vez finalizada la fase de identificación mediante imágenes tridimensionales de los granos de polen, puede existir un conjunto de puntos que no han podido ser identificados por el sistema. Para ello, se ha desarrollado un sencillo programa, “PolenGUT”, con una interfaz de usuario intuitiva que permite al experto palinólogo evaluar los resultados del sistema automático, permitiendo añadir o modificar un tipo o incluso eliminar un punto que no es grano de polen.

El interfaz presenta una tabla donde se muestra el identificador de grano detectado, su posición y su tipo. Se puede recorrer la tabla, grano a grano, y visualizar sobre la imagen digitalizada si el punto detectado corresponde a un grano, y por tanto identificarlo, o por el contrario eliminarlo. En el anexo A.4 se detalla el protocolo a emplear para verificar una lista de puntos de posibles granos de polen.

En la verificación de los granos de polen el sistema almacena información sobre las acciones realizadas por el experto, como la ratificación de que el objeto detectado es grano de polen con tipo correcto, la eliminación de objetos que no son granos de polen, la inserción del tipo de grano o su modificación. Toda la información almacenada se utilizará para la mejora del sistema en futuras versiones.

## 4.5. Resumen del capítulo

El sistema de detección e identificación de granos de polen presentado en este capítulo pretende ayudar al experto palinólogo en la tarea del recuento polínico, ya que solo se le necesita para la revisión de los resultados. Todo el proceso es automático y permite aplicar el protocolo de recuento polínico de la REA. Se compone de módulos que permiten desde la digitalización de todo el fragmento de 24 horas de captación para realizar el análisis de barridos horizontales hasta la identificación a alta resolución a partir de imágenes centradas en los posibles granos detectados de forma automática.

En la fase de detección a baja resolución, que es la etapa más compleja, se intentan localizar todos los objetos con características similares a los granos de polen. El prototipo actual tiene una sensibilidad del 77% y una precisión del 35%. Estos parámetros se mejorarán con el sistema a alta resolución 3D, que permite mejorar la calidad de las imágenes e identificar o descartar los granos de polen.

Se han desarrollado diferentes soluciones en las tres etapas del sistema: digitalización, detección e identificación, y se han utilizado dos tipos de resolución en las imágenes. A baja resolución se digitaliza la muestra de forma completa y se localizan objetos que sean posibles granos de polen. Si la información de la

imagen y su segmentación lo permiten, se puede identificar el tipo de grano a esta resolución. Para los objetos que no han podido ser identificados a baja resolución, se digitaliza a alta resolución imágenes centradas en los puntos detectados en la fase anterior. La digitalización se realiza a diferentes planos de enfoque, para tener información tridimensional del grano, y se analizan sus características para proceder a su identificación. Si no es posible una identificación automática del grano de polen a alta resolución, se dispone de un sistema de verificación que permite visualizar las imágenes del grano por los expertos para su identificación manual. Si las imágenes obtenidas del grano no fueran suficientemente buenas, el sistema permite la conexión remota con el microscopio óptico motorizado para adquirir nuevas imágenes centradas en el objeto de análisis.

Todas las modificaciones sobre la detección e identificación de granos realizadas sobre los resultados del sistema automático por los expertos son almacenadas, para mejorar el sistema en siguientes versiones. Pueden darse casos de objetos que no son granos de polen que sean identificados de forma incorrecta o que no se detecten. De esta forma, el sistema puede aprender de la historia de los granos de polen analizados y se posibilita su mejora de forma iterativa (metodología en espiral).

# Capítulo 5

## Resultados experimentales

El sistema descrito en el capítulo anterior es el resultado de los diferentes experimentos realizados en digitalización, detección e identificación de fragmentos diarios de muestras de contenido polínico aerotransportado.

### 5.1. Materiales

Se utilizan 12 fragmentos diarios de muestras de contenido polínico aerotransportado. Cada muestra diaria pertenece al día 15 de cada mes del año 2012 y procede de la succión de aire de un captador situado en Toledo. La muestra diaria pertenece a 24 horas de captación y su inicio se corresponde a las 12:00 pm horas de ese día.

Se dispone de un microscopio microscopio óptico y digital “NIKON 80i” con base motorizada “Prior H101 ProScanII” y el software NIS Elements Basic Research versión 4.11 de NIKON (ver figura 5.1). El software utilizado permite configurar el área a digitalizar y el nivel de enfoque de forma sencilla mediante su interfaz (ver figura 5.2). También permite su configuración mediante un lenguaje de macros (rutinas personalizadas) que es el método que se ha utilizado para la automatización de la digitalización.

Dado el gran tamaño de cada muestra diaria, se han analizado en este trabajo 4 líneas horizontales con el sistema automático de cada imagen, siguiendo las indicaciones de la REA, y se han comparado los resultados con el método tradicional y el método manual digital.

### 5.2. Métodos

Para el recuento automático de granos de polen a partir de una muestra procedente de la succión de aire son necesarias las siguientes tres fases: 1)

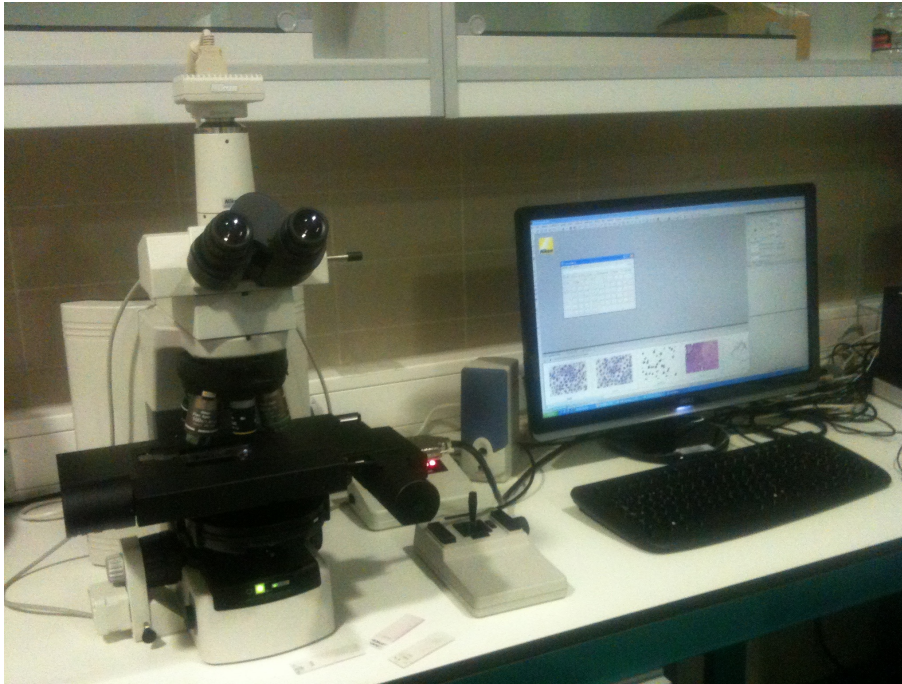


Figura 5.1: Fotografía del microscopio óptico y digital motorizado junto al PC con software NIS Elements

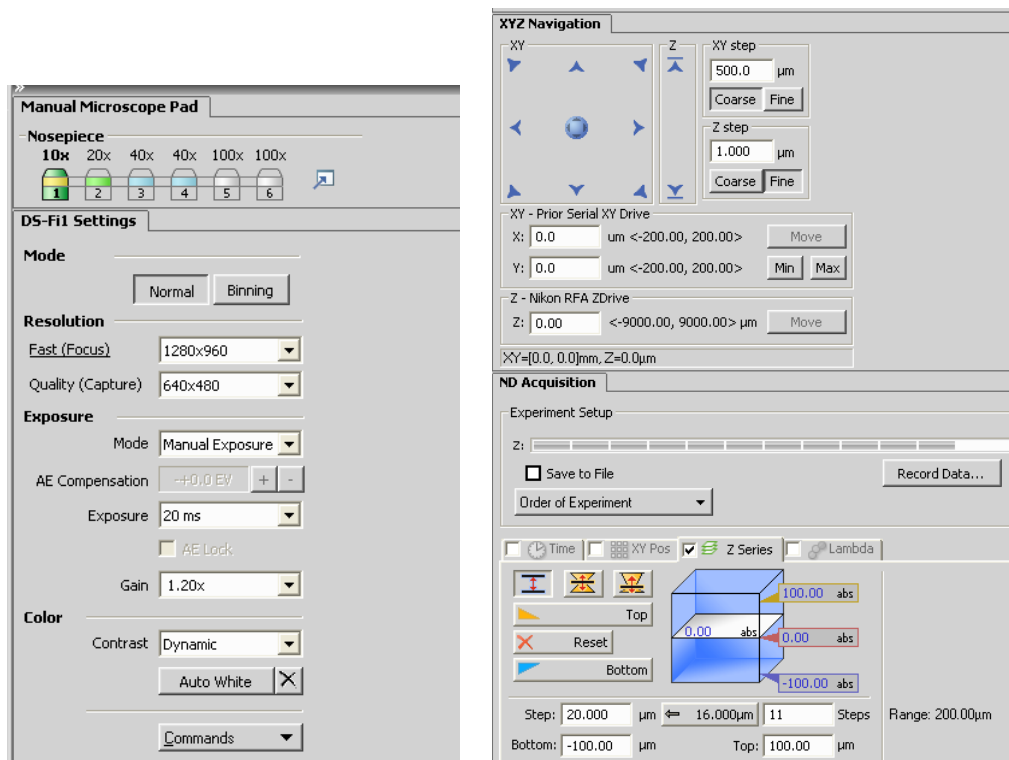


Figura 5.2: Módulos de configuración del software NIS Elements Basic Research

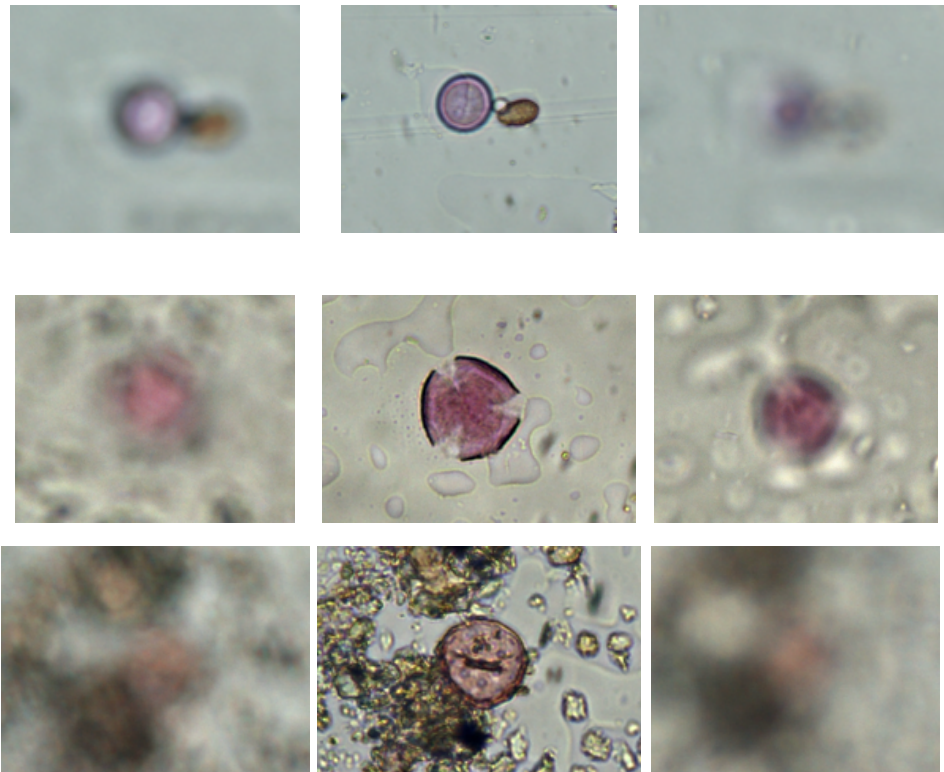


Figura 5.3: Granos de polen a diferentes valores de enfoque

Digitalización, 2) Detección y 3) Identificación. El sistema desarrollado que se presenta en el capítulo 4 se compone de las tres fases. A continuación, se comentan en detalle los pasos llevados a cabo en cada una de estas fases y se justifican las decisiones adoptadas en el desarrollo del sistema.

### 5.2.1. Digitalización

La digitalización de la muestra es la etapa más importante para obtener buenos resultados. Un enfoque adecuado de un grano de polen permite ver sus diferentes componentes, como la pared o las aperturas. Si se utiliza un valor de enfoque en el eje Z menor o mayor del óptimo pero dentro del grano, la imagen obtenida será la de un objeto rosáceo y difuminado. Si el enfoque se encuentra fuera del grano, la imagen solo contendrá información de fondo (ver figura 5.3).

En las primeras digitalizaciones, toda la muestra se escaneaba al mismo nivel del eje Z, partiendo de un enfoque óptimo en el centro de la muestra. Se observó que los extremos de la imagen estaban en la mayoría de los casos muy desenfocados por lo que se implementó una función dentro del software NIS Elements para estudiar el enfoque en una serie puntos de la muestra repartidos de forma equidistante. Se utilizan 49 puntos (ver fig. 4.6), elegidos entre el compromiso de eficiencia temporal

<u>Description</u>	<u>Return Values</u>
<pre>int StgFocusInRangeTwoPasses(     double FocusRange,     double CoarseStep,     double FineStep );</pre>	<pre>int</pre>
Performs two-pass autofocus with specified parameters.	<pre>1 The focus was successful 0 The focus failed -1 CoarseStep or FineStep is less than zero -2 FocusRange is less than zero -3 The focus failed (e.g. total white or black scene)</pre>
<u>Parameters</u>	
<pre>double FocusRange</pre> <p>Range in <math>\mu\text{m}</math>.</p>	
<pre>double CoarseStep</pre> <p>Coarse step in <math>\mu\text{m}</math> (first pass).</p>	
<pre>double FineStep</pre> <p>Fine step in <math>\mu\text{m}</math> (second pass).</p>	

Figura 5.4: Descripción de función de autoenfoco en software NIS

de la digitalización y el grado de enfoque obtenido.

Para la configuración de un método de digitalización mejorado de la muestra completa, ha sido necesario aprender el lenguaje de macros utilizado por el software NIS Elements Basic Research. Se ha creado una macro que permite digitalizar toda la muestra a 100 aumentos siguiendo unos pasos sencillos que se indican por pantalla. El experto solo debe colocar la muestra en el microscopio motorizado y enfocarla, en una posición central del fragmento, al inicio del proceso. El autoenfoco en cada uno de los puntos analizados se realiza en dos fases, con la función “intStgFocusInRangeTwoPasses” (ver descripción en figura 5.4) del software NIS Elements. En la primera fase se busca la imagen con mayor contraste en un rango de 600  $\mu\text{m}$  y salto de 25  $\mu\text{m}$ , y en la segunda, en un rango de 100  $\mu\text{m}$  sobre el óptimo obtenido en la primera fase y salto de 5  $\mu\text{m}$ .

Tras el procesamiento de la imagen de toda la muestra para detectar e identificar granos de polen de forma automática, se obtiene una lista con los puntos no identificados en la primera etapa. Se almacena en un fichero de datos las coordenadas en micrómetros (X, Y y Z) y el diámetro del objeto detectado que no ha sido identificado. Mediante otra macro desarrollada, se posiciona el microscopio motorizado en la posición indicada y se captura una pila de imágenes a 400 aumentos en función del diámetro calculado. En granos de tamaño pequeño se utiliza un salto de 1  $\mu\text{m}$ , en granos de tamaño medio salto de 5  $\mu\text{m}$  y en los grandes se utiliza un salto de 10  $\mu\text{m}$ . El aumento máximo utilizado es de 400 aumentos para que granos de polen grandes, como los tipos Pinus o Alnus ( $\sim 200 \mu\text{m}$ ), se digitalicen en una sola imagen.

La utilización de un enfoque automático de los granos puede fallar en algunos casos, debido a que aparezcan otras partículas en la imagen que produzcan un

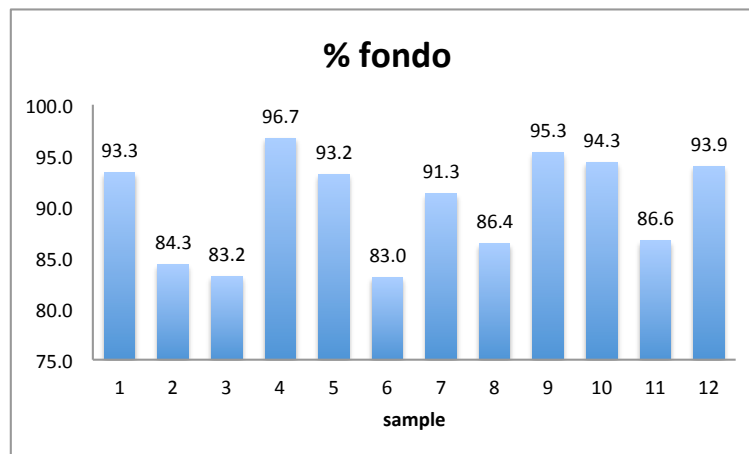


Figura 5.5: Porcentaje de fondo por muestra

enfoque óptimo en un valor Z fuera del grano. En estos casos, la pila de imágenes no contiene cortes de todo el grano por lo que su identificación se complica. Para estos casos, el sistema desarrollado permite la interacción remota con el microscopio para su posicionamiento y enfoque, de forma que se obtenga una nueva pila de imágenes que permita su identificación.

### 5.2.2. Detección

La detección de granos de polen en muestras diarias procedentes de la succión de aire es una tarea muy compleja, ya que 1) la imagen está compuesta por múltiples partículas, 2) el enfoque no es óptimo en toda la imagen, debido a su gran tamaño y 3) los granos de polen tienen un color y forma muy variado ya que el número de tipos contemplado es grande, en concreto, 38 tipos.

Una de las características básicas de las muestras de polen es la cantidad de fondo de la imagen, que da una primera idea de la composición de la muestra. Se puede conocer si la muestra es limpia o está compuesta de múltiples objetos, ya sean granos de polen u otro tipo de partículas como minerales o burbujas. En la figura 5.5 se muestra el porcentaje de fondo (suma de fondo claro y de objetos muy oscuros) sobre el total de cada línea analizada. Se observa que las muestras con un mayor contenido de partículas son las muestras pertenecientes al mes 2, 3, 6, 8 y 12.

La detección de granos de polen se ha realizado en varias fases. La primera fase consiste en la búsqueda de objetos con color y tamaño similar a los granos de polen, es decir, color rosáceo y de forma cuasi-circular. Mediante métodos de segmentación basados en píxeles y regiones se han desarrollado varios algoritmos para la detección automática de granos de polen, que analizan distintas propiedades visuales. Sus resultados representan distintas relaciones de compromiso entre detectar el mayor

número de granos de polen (verdaderos positivos) y el menor número de regiones que no correspondan a granos de polen (falsos positivos).

En las fases iniciales del algoritmo de detección se utilizaron umbrales fijos de color, tamaño y excentricidad para obtener posibles granos de polen. Su esquema de funcionamiento se presenta en la figura 5.6. Se han analizado cuatro barridos longitudinales de cada una de las muestras digitalizadas y se han comparado los objetos resultantes con los obtenidos por los expertos. Los resultados obtenidos con esta primera versión se muestran en la columna izquierda de la tabla 5.1. Se muestran los resultados por línea y se comparan con la anotación del experto que ha detectado un mayor número de granos. Se indican el número de granos detectados de forma manual por los expertos “N<sup>o</sup>Exp”, el número de granos detectados de forma automática “N<sup>o</sup>Auto” de cada algoritmo junto a el número de granos automáticos que tienen solape con los granos anotados por los expertos “N<sup>o</sup>Solape” y los no detectados en “N<sup>o</sup>noDet”. También se acumulan los resultados de todas las líneas (línea 0) para comparar los resultados a nivel de muestra. Se observó que existía un número alto de granos de polen no detectados, exactamente del 38.79 %, por lo que se ha mejorado el módulo de detección para minimizar este número.

El algoritmo actual, presentado en el capítulo 4, se basa en el análisis del espacio HSV en busca de objetos compactos de color rosáceos junto al análisis del espacio de color RGB con umbrales fijos. Su esquema de funcionamiento se presenta en la figura 4.7. Se ha analizado la misma superficie que con el algoritmo inicial y sus resultados se presentan en la columna derecha de la tabla 5.1. Para evaluar el algoritmo se utiliza el número de objetos detectados antes del análisis mediante la circularidad con la transformada de Hough.

La comparativa a nivel de muestra de los dos algoritmos de detección analizados se presenta en la tabla 5.2. El algoritmo inicial (algoritmo 1) detecta un menor número de objetos ya que tiene una configuración más restrictiva pero presenta una tasa alta de granos de polen no detectados, un 38.8 % frente al 6.02 % que presenta el algoritmo actual (algoritmo 2).

Se evalúan para cada algoritmo la sensibilidad (número de granos de polen detectados correctamente respecto al GS) 5.1, la precisión (número de granos de polen detectados respecto al total generado de forma automática) 5.2 y el f-score (relación ente sensibilidad y precisión) 5.3. El algoritmo 1 tiene una sensibilidad promedio del 56.16 % frente a un 89.17 % que presenta el algoritmo 2. Mientras que la precisión promedio en el algoritmo 1 es de 16.12 % frente al 3.40 % del algoritmo 2. Se considera mas importante la sensibilidad que la precisión , ya que los falsos granos de polen pueden ser analizados en etapas posteriores para intentar minimizar



Mes	Autor	Linea	NºExp	Algoritmo 1			Algoritmo 2		
				NºAuto	NºSolape	NºnoDet	NºAuto	NºSolape	NºnoDet
1	3	1	4	16	3	1	452	4	0
1	3	2	2	7	1	1	654	2	0
1	3	3	3	15	2	1	669	3	0
1	3	4	4	16	5	1	1111	4	0
1	3	0	13	54	11	4	2886	13	0
2	6	1	3	29	3	0	999	3	0
2	6	2	9	310	4	5	885	11	0
2	6	3	5	489	1	4	1281	4	1
2	6	4	3	204	0	3	1011	3	0
2	6	0	20	1032	8	12	4176	21	1
3	5	1	709	557	176	506	5334	697	84
3	5	2	335	1961	96	242	5757	287	64
3	5	3	446	4173	210	269	7330	452	17
3	5	4	276	736	67	209	6198	277	10
3	5	0	1766	7427	549	1226	24619	1713	175
4	14	1	18	100	8	8	359	31	2
4	14	2	19	119	14	10	355	39	1
4	14	3	16	112	14	8	419	38	0
4	14	4	19	117	23	6	503	33	3
4	14	0	72	448	59	32	1636	141	6
5	6	1	344	912	349	24	2867	328	7
5	6	2	485	1044	507	20	3315	456	3
5	6	3	413	955	457	21	3524	403	1
5	6	4	366	643	421	15	2757	351	3
5	6	0	1608	3554	1734	80	12463	1538	14
6	6	1	23	162	21	1	1546	23	0
6	6	2	73	498	81	5	2957	70	2
6	6	3	47	333	39	12	2171	47	0
6	6	4	59	222	31	28	1653	56	0
6	6	0	202	1215	172	46	8327	196	2
7	5	1	11	249	8	3	2765	11	0
7	5	2	9	254	9	1	2896	9	0
7	5	3	13	264	8	4	2837	10	0
7	5	4	11	200	11	4	3121	11	0
7	5	0	44	967	36	12	11619	41	0
8	4	1	11	69	12	1	2608	11	0
8	4	2	6	47	5	1	2506	5	0
8	4	3	2	73	3	0	2738	2	0
8	4	4	16	56	12	6	2543	15	1
8	4	0	35	245	32	8	10395	33	1
9	7	1	10	52	12	2	2548	10	0
9	7	2	7	79	8	1	3548	7	0
9	7	3	8	43	9	2	2678	8	0
9	7	4	6	49	5	1	3790	5	1
9	7	0	31	223	34	6	12564	30	1
10	3	1	9	328	3	6	2945	6	3
10	3	2	13	184	10	8	1917	11	2
10	3	3	19	274	6	11	1837	16	3
10	3	4	10	241	3	6	1954	6	3
10	3	0	51	1027	22	31	8653	39	11
11	5	1	6	105	4	4	1528	5	1
11	5	3	6	124	1	5	2023	6	0
11	5	4	15	103	4	13	2604	11	4
11	5	0	27	332	9	22	14808	61	16
12	5	1	9	31	1	8	446	7	2
12	5	2	10	17	3	7	499	9	1
12	5	3	22	19	5	19	339	16	5
12	5	4	9	16	2	7	267	8	1
12	5	0	50	83	11	41	1551	40	9

Tabla 5.1: Resultados de detección automática con imagen digitalizada a 100 aumentos

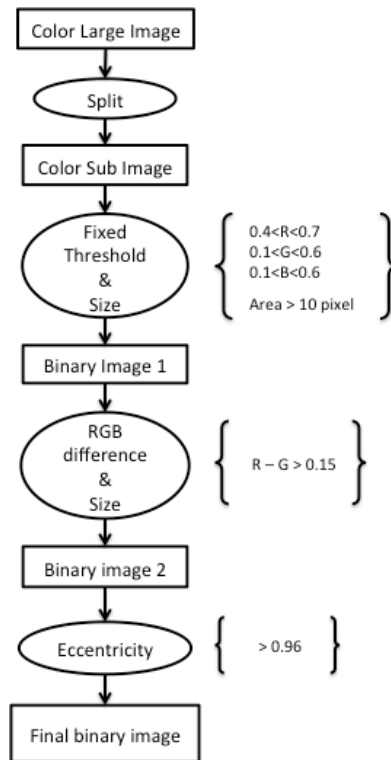


Figura 5.6: Esquema de trabajo del algoritmo de detección inicial

su presencia. Se ha calculado  $F_\beta$  con un valor  $\beta = 2$ , pero la importancia de la precisión en la detección de granos de polen puede ser más elevada.

El algoritmo 2 presenta muy buenos resultados ya que el objetivo del sistema completo es minimizar el número de falsos negativos, es decir, minimizar el número de granos no detectados.

$$Sensibilidad = \frac{Verdadero\ Positivo}{Verdadero\ Positivo + Falso\ Negativo} \quad (5.1)$$

$$Precisión = \frac{Verdadero\ Positivo}{Verdadero\ Positivo + Falso\ Positivo} \quad (5.2)$$

$$F_\beta = (1 + \beta^2) \frac{Precisión \times Sensibilidad}{(\beta^2 \times Precisión) + Sensibilidad} \quad (5.3)$$

El número de falsos positivos (objetos detectados que no son granos de polen) obtenido por el algoritmo actual es muy elevado, tiene una sobredetección del 9322 % sobre el número de granos de polen detectado por los expertos, por lo que se introducen nuevas etapas en el sistema de detección para reducirlo. Para ello, se analiza la segmentación obtenida mediante la transformada de Hough circular. La transformada de Hough es un método para la detección de contornos que utiliza

Mes	Autor	NºExp	Algoritmo 1			Algoritmo 2		
			NºAuto	NºSolape	NºnoDet	NºAuto	NºSolape	NºnoDet
1	3	13	54	11	4	2886	13	0
2	6	20	1032	8	12	4176	21	1
3	5	1766	7427	549	1226	24619	1713	175
4	14	72	448	59	32	1636	141	6
5	6	1608	3554	1734	80	12463	1538	14
6	6	202	1215	172	46	8327	196	2
7	5	44	967	36	12	11619	41	0
8	4	35	245	32	8	10395	33	1
9	7	31	223	34	6	12564	30	1
10	3	51	1027	22	31	8653	39	11
11	5	27	332	9	22	14808	61	16
12	5	50	83	11	41	1551	40	9
Total		3919	16607	2677	1520	113697	3866	236
Sensibilidad			56.16%			89.17%		
Precisión			16.12%			3.40%		
F2-Score			7.90%			41.00%		

Tabla 5.2: Comparativa entre algoritmos de detección automáticos

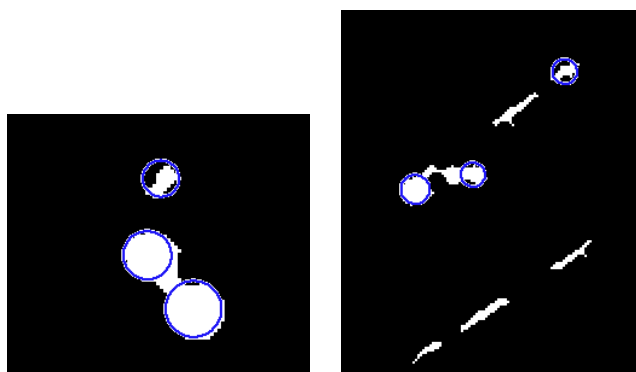


Figura 5.7: Resultados de circularidad en la segmentación con trasformada de Hough: Separación de objetos cercanos (izda.) y eliminación de objetos no circulares (dcha.)

conocimiento de la ecuación paramétrica del objeto que se busca en la imagen, que no se ve afectado por las lagunas existentes en la frontera del objeto ni por el ruido. La versión circular utiliza un acumulador de tres parámetros  $a$ ,  $b$  y  $r$ , ya que la ecuación paramétrica del círculo viene dado por  $(x - a)^2 - (y - b)^2 = r^2$ . En el caso de los granos de polen, se parte de un rango de posibles radios  $[Rmin Rmax]$  y se obtiene un lista de puntos con centro y radio del objeto detectado.

En función de la localización del centro y el radio, se descartan los objetos que no presentan circularidad y se dividen las segmentaciones de granos solapados o cercanos. En la figura 5.7 se presentan algunos ejemplos de los resultados de aplicar la trasformada de Hough circular sobre la segmentación obtenida en la etapa anterior.

Igual que en etapas anteriores, se aplica la trasformada circular de Hough a la segmentación obtenida en el sistema de detección. En la tabla 5.3 se muestran los resultados obtenidos en cada línea de barrido analizada por cada una de las muestras. Se han diferenciado los objetos en función del número de círculos en los que se segmenta la imagen. Se muestra la cantidad de blob que tienen un círculo, entre 2

y 4 círculos y más de 5 círculos. Si se descartaran todos los objetos segmentados que tienen más de un círculo, se obtendría una disminución del número de objetos a analizar del 24.98 % pero se pueden perder granos de polen. Por ahora, solo se descartan los objetos no circulares.

El número total de objetos detectados antes del análisis de la circularidad es de 106576 (ver tabla 5.4). Se descartan 17566 objetos no circulares y se dividen algunas segmentaciones en 9630 objetos circulares. Tras en análisis de circularidad se tienen 119666 puntos que son posibles granos de polen. Con la transformada de Hough circular se ha reducido el número de objetos un 15.39 % con una pérdida de verdaderos positivos de 5 %.

Tras el análisis de circularidad el número de falsos positivos sigue siendo elevado. Para tratar de minimizarlo, se han analizado 67 características de color, forma y textura en los objetos detectados. Se describe el objeto segmentado mediante características de forma y tamaño como el área, el eje mayor, el eje menor o la excentricidad. También se utilizan características de color, en espacio HSV y en espacio RGB, analizadas en distintas regiones de la imagen : todo el grano, su interior, la corteza o el exterior. Como medida de la variabilidad de los niveles de la imagen, se analizan características de textura como la entropía o de la matriz GLCM. En el anexo B se presenta de forma detallada las variables analizadas para entrenar un clasificador.

Se ha entrenado un clasificador de dos clases (0 para clase “No grano de polen” y 1 para clase “Si grano de polen”). Se han utilizado 123.135 instancias de las cuales 4941 pertenecen a granos de polen, ya que tiene solape con los granos de polen anotados por los expertos. Las 67 características de color, forma y textura se han realizado en anillos concéntricos en función del tamaño del grano detectado. Para el entrenamiento del clasificador basado en arboles de decisión se ha utilizado una muestra con el mismo número de instancias por cada clase y se han estudiado dos configuraciones. La primera con todas las variables estudiadas y la segunda con el conjunto de características mínimo que tiene un mayor ajuste a la clase de salida. La selección de atributos se ha realizado mediante una búsqueda de subconjuntos con un algoritmo genético con una población de 1000 y un número máximo de generaciones de 2000, obteniendo 16 variables (ver fig. C.3 ). En el anexo C se puede encontrar mayor información sobre los resultados de los clasificadores estudiados.

El clasificador que utiliza todas las características tiene una tasa de verdaderos positivos de 98.2 % y una tasa de falsos positivos de 9.1 % (ver fig. C.2), mientras que el clasificador que utiliza un subconjunto de 16 variables presenta una tasa de verdaderos positivos del 95.8 % y una tasa de falsos positivos del 8.4 % (ver fig. C.5).

Circular Hough transform													
Lugar	Mes	Autor	Línea	NºExperto	NºAuto blob	Nº auto circular blob	Nº blobs 1 seed	Nº blob 2 a 4 seed	Nº blob +5 seed	Blobs No circular	NºSolape	NºnoDet	
1	1	3	1	4	452	468	347	17	3	85	4	0	
1	1	3	2	2	654	717	483	47	6	118	2	0	
1	1	3	3	3	669	687	465	50	5	149	3	0	
1	1	3	4	4	1111	1087	821	59	6	225	4	0	
1	1	3	0	13	2886	2959	2116	173	20	577	13	0	
1	2	6	1	3	999	1003	752	70	2	175	3	0	
1	2	6	2	9	885	879	689	54	1	141	11	0	
1	2	6	3	5	1281	1329	931	95	4	251	4	1	
1	2	6	4	3	1011	1009	733	58	5	215	3	0	
1	2	6	0	20	4176	4220	3105	277	12	782	21	1	
1	3	5	1	709	5334	6023	3671	587	49	1027	705	101	
1	3	5	2	335	5757	6590	4154	603	40	960	280	69	
1	3	5	3	446	7330	9270	4807	950	122	1451	439	19	
1	3	5	4	276	6198	6770	3939	651	60	1548	273	11	
1	3	5	0	1766	24619	28653	16571	2791	271	4986	1697	200	
1	4	11	1	35	359	399	264	38	3	54	32	2	
1	4	11	2	41	355	405	256	50	2	47	39	1	
1	4	11	3	38	419	461	298	53	2	66	38	0	
1	4	11	4	38	503	579	366	61	5	71	33	3	
1	4	11	0	152	1636	1844	1184	202	12	238	142	6	
1	5	6	1	344	2867	3106	2195	263	19	390	327	7	
1	5	6	2	485	3315	3672	2499	317	29	470	461	3	
1	5	6	3	413	3524	3798	2695	319	22	488	407	1	
1	5	6	4	366	2757	2962	2129	232	18	378	359	4	
1	5	6	0	1608	12463	13538	9518	1131	88	1726	1554	15	
1	6	5	1	23	1546	1599	1180	102	7	257	23	0	
1	6	5	2	73	2957	3063	2193	215	17	532	71	2	
1	6	5	3	47	2171	2226	1593	152	12	414	47	0	
1	6	5	4	59	1653	1717	1212	131	10	300	57	0	
1	6	5	0	202	8327	8605	6178	600	46	1503	198	2	
1	7	4	1	11	2765	2815	2082	201	13	469	11	0	
1	7	4	2	9	2896	2986	2256	215	5	420	9	0	
1	7	4	3	13	2837	3089	2161	249	17	410	11	0	
1	7	4	4	11	3121	3322	2358	258	16	489	11	0	
1	7	4	0	44	11619	12212	8857	923	51	1788	42	0	
1	8	4	1	11	2608	3577	1846	313	68	381	11	0	
1	8	4	2	6	2506	3774	1767	353	66	320	5	0	
1	8	4	3	2	2738	4115	1965	313	55	405	2	0	
1	8	4	4	16	2543	4143	1809	308	72	354	15	1	
1	8	4	0	35	10395	15609	7387	1287	261	1460	33	1	
1	9	6	1	10	2548	2536	1982	166	4	396	10	0	
1	9	6	2	7	3548	3591	2804	226	13	505	7	0	
1	9	6	3	8	2678	2739	2146	169	10	353	8	0	
1	9	6	4	6	3790	3911	2976	293	14	507	5	1	
1	9	6	0	31	12564	12777	9908	854	41	1761	30	1	
1	10	3	1	9	2945	3113	2264	216	16	449	6	3	
1	10	3	2	13	1917	2047	1457	161	10	289	11	2	
1	10	3	3	19	1837	2008	1406	150	18	263	16	3	
1	10	3	4	10	1954	2083	1457	181	10	306	6	3	
1	10	3	0	51	8653	9251	6584	708	54	1307	39	11	
1	11	5	1	6	1528	1641	1172	108	14	234	5	1	
1	11	5	2	6	1528	1641	1172	108	14	234	0	0	
1	11	5	3	6	2023	2175	1531	171	12	309	6	0	
1	11	5	4	15	2604	2902	1939	279	15	371	11	4	
1	11	5	0	33	7683	8359	5814	666	55	1148	61	16	
1	12	4	1	9	446	444	326	28	2	90	7	2	
1	12	4	2	10	499	536	372	29	6	92	9	1	
1	12	4	3	22	339	379	252	26	4	57	16	5	
1	12	4	4	9	267	280	194	20	2	51	8	1	
1	12	4	0	50	1551	1639	1144	103	14	290	40	9	

Tabla 5.3: Resultados obtenidos tras analizar la segmentación de granos de polen con transformada de Hough circular por línea

Mes	NºExperto	NºAuto blob	Circularidad con Transformada de Hough							Clasificador		
			Nº blobs solape	Nº blobs no detectados	Nº auto circular blob	Nº blobs 1 seed	Nº blob 2a4 seed	Nº blob +5seed	Blobs No circular	Solape con GS	TP	FP
1	13	2886	13	0	2959	2116	173	20	577	13	12	150
2	20	4176	21	1	4220	3105	277	12	782	19	19	357
3	1766	24619	1697	200	28653	16571	2791	271	4986	1463	1440	6398
4	152	1636	142	6	1844	1184	202	12	238	122	120	164
5	1608	12463	1554	15	13538	9518	1131	88	1726	1449	1389	401
6	202	8327	198	2	8605	6178	600	46	1503	197	192	1020
7	44	11619	42	0	12212	8857	923	51	1788	41	39	584
8	35	10395	33	1	15609	7387	1287	261	1460	33	27	186
9	31	12564	30	1	12777	9908	854	41	1761	30	24	201
10	51	8653	39	11	9251	6584	708	54	1307	38	23	216
11	27	7683	61	16	8359	5814	666	55	1148	22	22	1158
12	50	1551	40	9	1639	1144	103	14	290	42	42	184
<b>Total</b>	<b>3999</b>	<b>106572</b>	<b>3870</b>	<b>262</b>	<b>119666</b>	<b>78366</b>	<b>9715</b>	<b>925</b>	<b>17566</b>	<b>3469</b>	<b>3349</b>	<b>11019</b>

Tabla 5.4: Resultados obtenidos tras analizar la segmentación de granos de polen con transformada de Hough circular por muestra

Se decide utilizar el clasificador que utiliza un subconjunto de los datos, ya que permite disminuir el tiempo de computo de las características, presenta una mejor tasa de FP aunque exista una pérdida de verdaderos granos de polen ligeramente superior que en el clasificador con todas las características.

El sistema a baja resolución presenta muy buenos resultados. El número de falsos positivos se reduce en gran medida tras el análisis de la circularidad mediante la transformada de Hough y tras el clasificador. Se disminuye el número de muestras a analizar un 84% con una pérdida de verdaderos granos de polen de un 5.2%. En la siguiente etapa de análisis se va a mejorar aún más la tasa de error de falsos positivos a partir de la mejora de la resolución de la imagen de los granos de polen.

### 5.2.3. Identificación

Debido a la complejidad de la detección de granos de polen a baja resolución, no se ha podido desarrollar aún los módulos de identificación en el sistema propuesto, por lo que se encuentran en etapas tempranas. Se ha implementado un clasificador con las características visuales a 100 aumentos de los granos de polen que aparecen en al menos 50 ocasiones. De los 38 tipos previstos, solo hay 7 tipos de granos que tienen una presencia suficientemente elevada como para poder ser utilizados en el entrenamiento de un clasificador. Son los tipos 13, 22, 26, 28, 29, 30 y 31. Se obtienen los atributos que tienen más peso en la variable de salida y se obtienen 19 variables significativas (VerdeMedio, AzulMedio, HMedio, SMedio, SDesv, A1RojoMedio, A1AzulMedio, A1HMedio, A1HDesv, A1SMedio, A2VerdeDesv, A2AzulMedio, A2SMedio, A2SDesv, A3RojoMedio, A3HMedio, A1Contrast, A1Energy, A3Homogeneity). Se entrena un árbol de decisión con 3363 instancias y 8 clases y se obtiene un árbol de decisión con 268 hojas y un tamaño de árbol de 535 nodos. En la figura 5.8 se muestra el resumen de resultados tras evaluar el clasificador. El porcentaje de instancias mal clasificadas es elevado por lo que es necesario mejorar el modelo.

## 5.3. Resumen del capítulo

Los experimentos realizados en este capítulo han sido los necesarios para desarrollar el sistema de detección e identificación de muestras polínicas procedentes de la succión de aire descrito en el capítulo 4. Para ello se han realizado experimentos en las diferentes fases en que se compone el sistema (digitalización, detección e identificación). Se han utilizado 12 muestras procedentes de un captador situado en Toledo, cada una procedente de un mes del año 2012. Se han realizado experimentos

```

Correctly Classified Instances      2551          75.8549 %
Incorrectly Classified Instances    812           24.1451 %
Kappa statistic                    0.6813
Mean absolute error                 0.067
Root mean squared error             0.2341
Relative absolute error             35.2388 %
Root relative squared error        75.9214 %
Total Number of Instances          3363

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      0.44     0.034   0.435     0.44    0.438     0.693     0
      0.751    0.061   0.775     0.751   0.763     0.871    13
      0.678    0.006   0.678     0.678   0.678     0.886    22
      0.669    0.01    0.742     0.669   0.704     0.837    26
      0.2       0.014   0.203     0.2     0.202     0.649    28
      0.331    0.024   0.379     0.331   0.353     0.65     29
      0.784    0.108   0.757     0.784   0.77     0.83     30
      0.906    0.048   0.892     0.906   0.899     0.937    31
Weighted Avg.  0.759    0.065    0.756    0.759    0.757    0.854

=== Confusion Matrix ===

  a  b  c  d  e  f  g  h  <-- classified as
84 11  8  7  3 15 23 40 | a = 0
15 553 1  3  9  9 144  2 | b = 13
 8  3 40  0  0  2  5  1 | c = 22
 7  4  1 89  0  8 13 11 | d = 26
 0  6  0  1 12  2 34  5 | e = 28
26  9  4  5  2 47 16 33 | f = 29
19 122 3  7 29 18 793 21 | g = 30
34  6  2  8  4 23 20 933 | h = 31

```

Figura 5.8: Resultados de entrenamiento y test de árbol de decisión para identificación polínica a 100 aumentos

en la fase de digitalización y en la de detección a baja resolución, y se espera mejorar el sistema con experimentos en la fase de identificación.

La digitalización de la muestra diaria es fundamental para obtener buenos resultados en la siguientes fases. Por ello, se utiliza una malla de 49 puntos para analizar el enfoque y proceder a la digitalización con el plano de menor error medio de enfoque.

Debido a que el objetivo del sistema es el recuento polínico, los módulos desarrollados intentar de ofrecer buenos resultados minimizando el número de granos no detectados. Para ello, se ha desarrollado un módulo de detección que busca objetos de color rosáceo con restricciones de tamaño y forma poco exigentes, que genera un número elevado de posibles granos de polen. Mediante el uso de la transformada de Hough circular y el análisis de características locales y regionales del grano se reduce el número de falsos negativos un 91 % mediante la utilización de un clasificador basado en arboles de decisión.





# Capítulo 6

## Conclusiones y trabajo futuro

### 6.1. Discusión

Se ha desarrollado un sistema para la automatización de la detección e identificación de granos de polen mediante la digitalización de muestras polínicas, que permite ahorrar tiempo y realizar un análisis más exhaustivo de los granos de polen a los expertos del dominio. Además, almacena la información de la fecha, la posición y el tipo de grano junto a sus características visuales, que permite realizar diferentes tipos de estudios.

El análisis de las muestras a distintos aumentos permite trabajar con toda la imagen, permitiendo la detección de posibles granos de polen para a continuación obtener imágenes de mayor calidad e información a distintos planos de enfoque, localizadas en las zonas de interés. Esta metodología permite un gran ahorro de tiempo de adquisición y de cantidad de memoria, ya que se digitaliza toda la muestra a baja resolución y solo se adquieren imágenes de alta resolución en los puntos donde se detecta un posible grano de polen.

El uso de herramientas como la desarrollada junto a algoritmos automáticos de detección de granos permite reducir la carga de trabajo del experto, ya que solo tendrán que supervisar los casos dudosos. Además, el almacenamiento digital de la información de los granos de polen permitirá realizar en un futuro otros tipos de análisis y usos, tanto en docencia como en investigación, ya sean estudios temporales, análisis de la variabilidad intraexperto, etc.

Uno de los problemas encontrados ha sido la digitalización de la muestra, donde han surgido infinidad de problemas con el sistema hardware y software suministrado por NIKON, lo que ha provocado una carga de trabajo adicional. Por otro lado, también ha sido necesario dedicar un tiempo extra al enfoque correcto de la muestra, ya que un enfoque inadecuado de toda la cinta provoca que los algoritmos de

detección no localicen correctamente los posibles granos de polen. Para solventarlos se han desarrollado diferentes funciones en MATLAB® como el análisis del plano de menor error en el enfoque.

En cuanto a los algoritmos analizados, el porcentaje de sobredetección es menor en el algoritmo inicial pero el porcentaje de granos de polen detectados es mucho mayor en el algoritmo actual. Por ello, se utilizarán los resultados del algoritmo actual en una primera fase de detección junto al análisis de circularidad y de características de la imagen para reducir el número de se de falsos positivos. De igual forma que se utiliza en (Rincón et al., 2011).

En próximas versiones del sistema se desarrollaran los módulos de identificación de granos de polen tanto a 100 aumentos como a 400 aumentos y se mejorarán los módulos ya desarrollados.

## 6.2. Conclusiones

Los objetivos planteados al inicio de este trabajo han sido cumplidos con éxito ya que se ha desarrollado e implementado un sistema que permite mejorar la tarea de detección e identificación de granos de polen mediante la utilización de herramientas hardware y software interconectadas, es decir, mediante un microscopio digital controlado por un ordenador conectado a internet. El sistema desarrollado permite digitalizar un fragmento de 24 horas de captación polínica con un plano de enfoque variable que se ajusta al contenido de la imagen, ya que se analiza una malla de puntos previo a la digitalización.

Gracias al almacenamiento digital de la información, la detección y recuento polínico puede realizarse en cualquier ordenador de forma que el experto palinólogo no tiene que desplazarse hasta el microscopio, que suele encontrarse en salas frías para preservar la conservación de las preparaciones.

El sistema desarrollado puede utilizarse de forma completamente manual, tanto a 100 aumentos como a 400 aumentos, para detectar e identificar los diferentes granos de polen presentes en la muestra. Además, el programa interfaz desarrollado tiene facilidades para realizar anotaciones en las imágenes siguiendo las directrices de la REA, que aconsejan la realización de 4 barridos horizontales a lo largo de toda el fragmento. Se ha evaluado el sistema de forma manual y se ha obtenido un porcentaje de identificación en la primera fase del 44 % de los granos detectados y de un 74 % en la segunda.

Para ayudar aún más a los expertos en la tarea del recuento polínico se ha comenzado a desarrollar un sistema automático para la detección e identificación

de granos de polen. Este sistema se compone de un módulo de digitalización, un módulo de detección y un módulo de identificación. Debido a la complejidad que supone la detección automática de granos de polen, ha dado tiempo a desarrollar los sistemas de digitalización y de detección mientras que la fase de identificación aún se encuentra en una fase temprana.

El módulo de digitalización ofrece muy buenos resultados. Antes de la utilización del clasificador en el módulo de detección a baja resolución se obtiene una tasa baja de falsos negativos (menos del 3%) y una tasa alta de verdaderos positivos (96%) pero con un número muy alto de falsos positivos. Para reducir estos últimos, se utiliza un clasificador que reduce en gran medida los falsos positivos (84%) a costa de aumentar los falsos negativos a un 6%.

En las próximas versiones se pretende mejorar el clasificador a baja resolución para evitar la pérdida de verdaderos granos de polen y junto al sistema a alta resolución, se espera obtener unos resultados mejorados que permitan plantear modificar el protocolo actual en la tarea de recuento polínico para introducir sistemas inteligentes que imiten la labor del experto.

### 6.3. Aportaciones a la investigación

Las contribuciones realizadas por este trabajo de investigación se sintetizan en los siguientes puntos:

1. Se propone una nueva metodología en el proceso de recuento de granos de polen a partir de la utilización de muestras digitalizadas.
2. Se presenta un sistema computacional que analiza una muestra procedente de la succión de aire a dos resoluciones. A baja resolución se realiza la detección y a alta resolución la identificación.
3. Se propone la utilización de las nuevas tecnologías para el acceso remoto a los dispositivos, permitiendo analizar una muestra polínica a distancia.
4. Se crea una base de imágenes de muestras diarias que permite compartir datos y realizar otros estudios.

El sistema desarrollado se ha implementado en el Instituto de Ciencias Ambientales de la Universidad de Castilla-La Mancha, que es la sede de la Red de Aerobiología de Castilla-La Mancha (AEROCAM<sup>1</sup>) y está dirigido desde el área de Botánica

---

<sup>1</sup><http://aerocam.uclm.es>

de la UCLM. Con este sistema se pueden almacenar las muestras polínicas diarias digitalmente y analizar su contenido de forma automática para llevar a cabo nuevos proyectos de investigación, con características como el tamaño o el color de los granos de polen.

Los resultados de este desarrollo fueron presentados en el segundo congreso internacional APLE-APLF “Pollen biotechnology, diversity & function in changing environment” celebrado en Madrid del 17 al 20 de Septiembre de 2013. Los comentarios fueron muy positivos, pues se retoma una tarea que había sido comenzada por varios organizadores de este congreso.

## 6.4. Trabajo futuro

En próximas versiones del prototipo desarrollado se pretende mejorar los módulos actuales de digitalización y detección. Además, se comenzara el desarrollo del módulo de identificación que permita reconocer los granos de polen dentro de los 38 tipos previstos.

Se incrementará el número de muestras digitalizadas de forma que una mayor cantidad de datos etiquetados permita generalizar modelos de los granos de polen más robustos. Se pretende analizar la distribución temporal (diaria, mensual o anual) y espacial (en función de la localización del captador) de los distintos tipos de granos de polen para mejorar los modelos de identificación.

A partir de las muestras digitalizadas se puede generar una base de datos que contenga información para llevar a cabo nuevos estudios, como el análisis de la contaminación atmosférica o el efecto de las condiciones meteorológicas en la aparición de los tipos polínicos.

La utilización de las anotaciones almacenadas por el sistema desarrollado es muy variada, permitiendo realizar estudios de predicción de concentraciones de polen para uso medico como control de alergias o predicción de concentraciones para uso agrícola como el control de plagas, la predicción de cosechas o el control de regadíos.

# Apéndice A

## Protocolo de digitalización y análisis

Este protocolo indica los pasos a seguir para la digitalización y análisis de una muestra polínica con un microscopio NIKON 80i y el software NIS Elements Basics Research versión 4.11 de NIKON ubicados en el laboratorio 0.25 del Instituto de Ciencias Ambientales de la Universidad de Castilla-La Mancha, Toledo (España).

### A.1. Digitalización de la muestra

A continuación se enumeran los pasos que se deben seguir para digitalizar una muestra:

1. Colocar la muestra en la platina verificando que queda bien ajustada en los bordes.
2. Comprobar que el zoom utilizado en el microscopio es de 10 aumentos.
3. Posicionar el objetivo con el software NIS Elements en el centro de la muestra, alrededor de  $X = 50000 \mu\text{m}$  e  $Y = 35200 \mu\text{m}$ .
4. Visualizar en el software NIS la imagen pulsando el botón "Live".
5. Si la imagen se ve oscura, ajustar valores de "Exposure".
6. Verificar que "Gain" tiene el valor 1.2x.
7. Pulsar botón "Autowhite" para mejorar la visualización de la imagen.
8. Ajustar el enfoque hasta ver nítida la imagen. Se puede hacer manualmente con la ruleta gris del lado izquierdo del microscopio o con "REMOTE FOCUS ACCESSORY". Si no hay información suficiente mover la muestra a un punto cercano y enfocar.

9. Resetear “REMOTE FOCUS ACCESSORY” para iniciar la Z a 0.0  $\mu\text{m}$ , pulsando el botón negro que tiene al lado de la pantalla.
10. Si la imagen se ve muy oscura o demasiado clara, ajustar la posición de la lampara con la ruleta negra que hay a la izquierda del microscopio. Si se desenfoca la imagen volver al paso 8.
11. Verificar en el panel derecho de NIS Elements que el método de interpolación en el módulo Focus Surface es “Smooth”. Si no lo es cambiarlo.
12. Ejecutar la macro que captura todo el portaobjetos. Ir a Macros > Run from... y seleccionar el fichero  
*C:\Proyecto\_Polen\_UCLM\_UNED\Macros\ Scan\_Npoints\_ & \_All\_Tape.mac.*

La macro diseñada para la captura de toda la muestra puede capturar toda la muestra utilizando unos valores por defecto o puede capturar la muestra indicando los cuatro bordes de la cinta y la posición de las cuatro líneas negras. Aparecerán una serie de ventanas donde introducir información o responder preguntas según el tipo de captura deseada. A continuación se explican en detalle los pasos:

1. Aparece una ventana donde hay que introducir el nombre y fecha de la muestra (Ej: Toledo\_15-03-2012).
2. Responder a la pregunta: “¿Quiere buscar los limites de la cinta y las posiciones de lectura manual?” Aparecen cuatro botones con las siguientes opciones:
  - a) Yes: Se va a seleccionar los bordes de la cinta y los puntos de la líneas negras.
    - 1) Recorrer con el joystick para ver con NIS la posición mínima del borde inferior, la mínima del lado derecho y las máximas del borde superior y del borde izquierdo (las cuatro esquinas de la cinta). Pulsar el botón OK para continuar.
    - 2) Mover el joystick para ver con NIS cada una de las cuatro líneas negras. La línea 1 es la superior según se mira la muestra. Pulsar el botón OK para continuar.
  - b) No: Se pregunta al usuario que introduzca los limites de la cinta y la posición de las líneas. Por defecto tiene configurados unos valores que contienen la cinta y estiman las 4 líneas.
  - c) From file: No implementado

- d) Exit: Salir sin capturar
3. Responder a la pregunta: ¿Quiere cambiar los parámetros configurados en la captura?
- a) Yes: Aparecen ventanas para insertar los valores de las configuración a modificar.
  - b) NO.
4. Enfocar en el centro de la captura: Se posiciona automáticamente la platina en el centro de los límites introducidos y se pide de nuevo enfocar la muestra y resetear “REMOTE FOCUS ACCESORY” para iniciar la Z a 0.0  $\mu\text{m}$ , pulsando el botón negro que tiene al lado de la pantalla. Pulse Ok para continuar. (NOTA: Punto 24 debería estar enfocado en Z=0 aprox )
5. Responder a la pregunta: “¿Quiere capturar puntos de referencia?”
- a) YES: Se van a enfocar 49 puntos del interior de la cinta. El proceso dura unos 30 min y se obtiene la posición y el enfoque de cada punto que se introduce en la lista de puntos del módulo Focus Surface.
    - 1) Automatic: De forma automática se autoenfocan los puntos. (No funciona bien ya que hace una búsqueda de bordes)
    - 2) Manual: El enfoque se realiza de forma manual girando la ruleta de “REMOTE FOCUS ACCESORY” hasta una visualización por pantalla correcta. Pulsar OK para pasar al siguiente punto. (NOTA: Una vez validado estudiaremos como bajar este número de puntos)
  - b) NO. Responder a la pregunta: “¿Quiere utilizar FocusSurface?”
    - 1) YES: Utilizará en la captura de todo el porta la lista de puntos con cada valor Z previamente almacenado en el módulo Focus Surface.
    - 2) NO: Se pregunta al usuario que valor de enfoque, Z, se desea utilizar en la captura de todo el porta.
6. Responder a la pregunta: ¿Es correcta la lista de puntos y de valores de Focus Surface? Esto es una vía de escape para no tener que empezar de nuevo a enfocar todos los puntos si hay un error en alguno.
- a) YES. Continúa el proceso

b) NO. Finaliza la macro para que se verifique la lista de puntos y sus enfoques. En el módulo Focus Surface se recorrer la lista de puntos para verificar Z (activar selector de mostrar la imagen seleccionada de la lista). Si algún punto necesita redefinir el valor de Z pulsar el botón Z↵. Lanzar de nuevo la macro con el mismo nombre pero seleccionar No en la pregunta 2 e introducir los valores (sin decimales) que se encuentran en el fichero \*info0.txt del directorio de la muestra (Left, Right, Top y Bottom). Y responder NO a la pregunta 3, pulsar OK en el punto 4 y responder NO en pregunta 5. Por ultimo responder YES a utilizar FocusSurface y YES en esta pregunta, la número 6 para continuar el proceso.

7. Comienza la captura de toda la muestra. Este proceso dura 1 hora aproximadamente.

Al finalizar la captura, en el directorio

*C:\Proyecto\_Polen\_UCLM\_UNED\Imágenes\\*\*NOMBRE\_MUESTRA\*\**

se tiene la imagen total y un fichero de información sobre la configuración de la captura. Si se han adquirido puntos de referencia se encontrará 49 imágenes.

## A.2. Análisis de la muestra

Para el análisis de la muestra se ha desarrollado un programa con el software Matlab denominado “PolenGUP”. Con este programa se analiza la imagen capturada a 10x con el objetivo de encontrar y clasificar los diferentes tipos de polen que contiene la muestra.

El interfaz es sencillo y permite cargar la imagen de una muestra para editar las posiciones y el tipo de polen (ver figura A.1). Los pasos a seguir para analizar la imagen son los siguientes:

1. Seleccionar el directorio donde se encuentra la imagen junto a los puntos de referencia y el fichero de configuración.
2. Si es la primera vez que se utiliza el directorio:
  - a) Se analizan los puntos de referencia para calcular el valor de  $\mu\text{m} / \text{px}$  utilizado, la posición inicial en  $\mu\text{m}$  y obtener los valores de enfoque utilizados (Z) y sus posiciones (X,Y).
  - b) Se ajustan los bordes de la muestra. Se muestra una imagen del borde derecho y del borde izquierdo de la muestra para que el usuario marque las esquinas de la cinta.



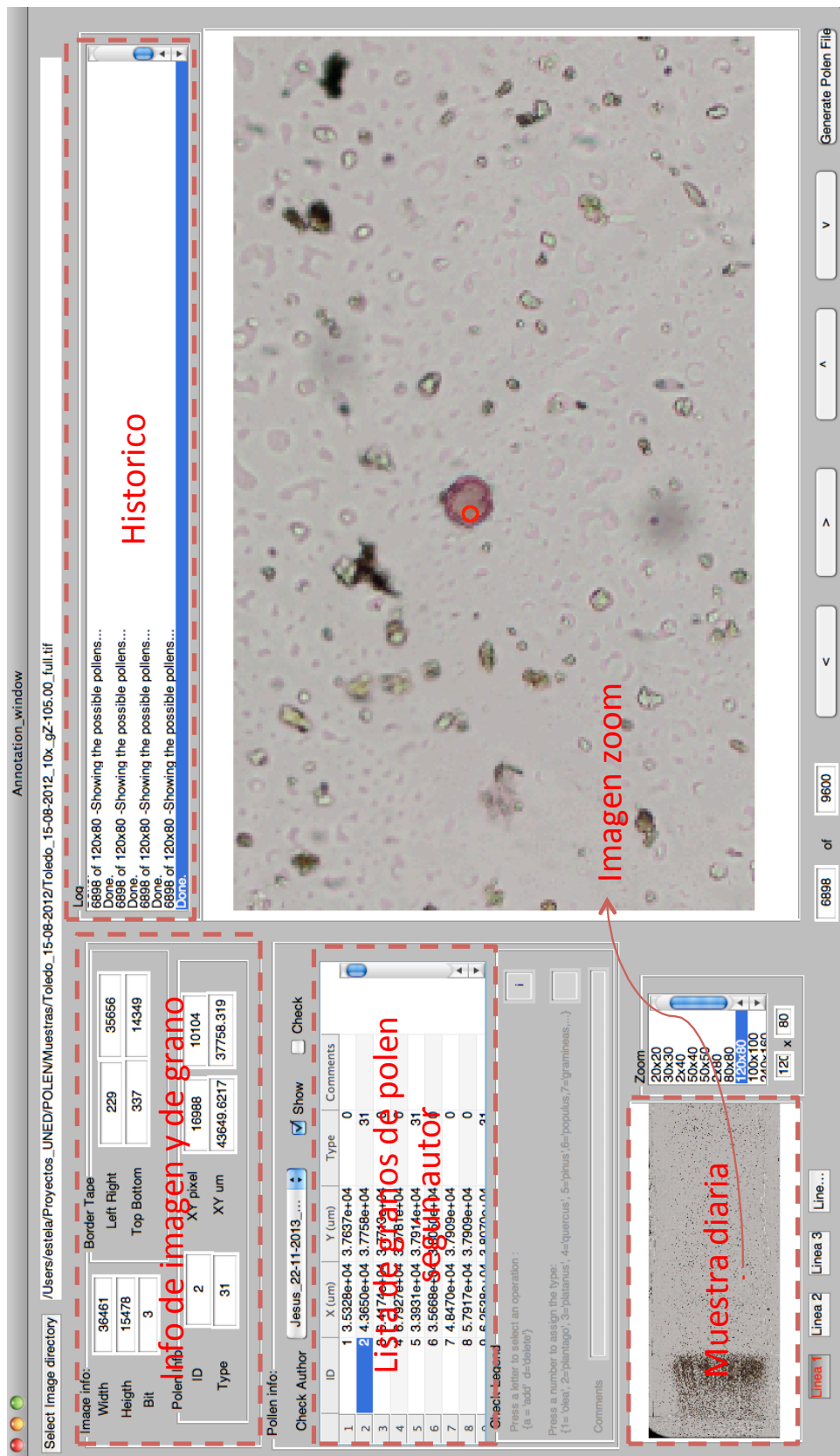


Figura A.1: Interfaz de PolenGUI

- c) Si no se han adquirido las posiciones de las 4 líneas en la fase de configuración de la captura se pide al usuario que las añada. ¿Quiere seleccionar las posiciones de las cuatro líneas negras?
    - 1) Pulse SI, cuando la imagen se ha tomado con los valores por defecto y seleccione las cuatro posiciones.
    - 2) Pulse No, y se calcularán de forma estimada en función del tamaño de la cinta.
  - d) Se analiza la imagen para localizar y clasificar los distintos tipos de polen que aparecen en la muestra.
3. Se carga la información sobre la posición y tipo del polen de la muestra analizada (como mínimo siempre se tiene la información del proceso automático).
  4. Ajustar zoom. El zoom de la imagen se inicia con unas divisiones de 120 x 80. Existen dos formas de ajustarlo:
    - a) Lista de divisiones. Seleccione de la lista la configuración mas adecuada para ver correctamente la muestra en la ventana de la izquierda.
    - b) Valores de las divisiones en horizontal y en vertical. Edite el valor para obtener la visualización adecuada.
  5. Seleccionar posición. El rectángulo rojo de la imagen inferior derecha orienta del tamaño que se muestra y de la posición. Existen dos formas de variar la posición:
    - a) Botones derecha e izquierda: Estos botones aumentan o disminuyen la posición del rectángulo rojo según el número de veces pulsado.
    - b) Botones arriba y abajo: Estos botones aumentan o disminuyen la posición del rectángulo rojo según el número de veces pulsado.
    - c) Pinchando al rectángulo rojo: Al seleccionar el rectángulo rojo este cambia a un tipo de línea discontinua para a continuación desplazarse al punto que se haya pinchado con el ratón.

NOTA: Los botones de Linea1 a Linea4 sirven para posicionar el rectángulo de visualización al inicio de la muestra donde se encuentra una marca negra en la platina para realizar un barrido horizontal.

NOTA: La anchura de la cinta en vertical son 20 mm, por lo que para recorrer una anchura de 0.5 mm y que se vea bien el contenido se utiliza un zoom de 120 x 80. Se han de recorrer 240 rectángulos, que son dos líneas en vertical. (OJO: Al cambiar el zoom se deselecciona la línea X. Para anotar compruebe que esta seleccionada. Si no, pulse en el botón de línea X).

6. Mostrar polen. Si se encuentra activa la casilla “SHOW” se muestra en la imagen de la derecha, mediante cruces azules, la situación de los polenes encontrados según el autor que se encuentre cargado.

Se puede recorrer la lista de polen anotado recorriendo la tabla de “Pollen info”. El polen seleccionado se muestra con un círculo rojo.

NOTA: Con la función “información” (pulsar tecla “i”) se puede conocer mas información pinchando en el punto semilla del polen.

7. Editar muestra. Si se encuentra activa la casilla “CHECK” se pueden añadir y borrar de la lista el polen encontrado. Existen dos modos, uno para la anotación y otro para la verificación de la anotación.

- a) Modo anotación: Añadir el nombre del autor (Ej: Maria\_12-03-2013) y seleccionar “NO” en la ventana que aparece.

NOTA: Si se escribe de nuevo un nombre que ya existe saldrá una ventana para elegir si continuar (“Continue”) con la lista de anotación, eliminar (“Delete”) la lista o cancelar (“Cancel”). Esto sirve para retomar el trabajo de anotación o revisar lo que se ha hecho por el mismo autor de forma manual con desplazamiento por la imagen.

- 1) Posicionar la imagen en la zona a anotar: Si se va a anotar alguna línea pulsar botón de línea correspondiente para ir a su inicio y recorrerla el número de rectángulos en función del zoom utilizado. Mientras esté pulsado (letras rojas) se impide salir del área de análisis de 500  $\mu\text{m}$  con los botones de desplazamiento
- 2) Añadir polen (ADD): 1º insertar el tipo de polen, en la caja de texto introducir el número del tipo de polen, pulsar ENTER. Después pulsar la tecla ‘a’ de añadir polen, hacer click en el polen correspondiente. Aparecerá un asterisco de tonos azules. Si se ha añadido un punto erróneo se puede borrar el punto con el botón de

ID	Nombre	ID	Nombre
0	No identificado	20	Ligustrum
1	Acer	21	Mercurialis
2	Alnus	22	Moraceae
3	Apiaceae	23	Myrtaceae
4	Artemisia	24	Olea
5	Betula	25	Palmae
6	Brassicaceae	26	Pinus
7	Cannabinaceae	27	Plantago
8	Castanea	28	Platanus
9	Casuarina	29	Poaceae
10	Cedros	30	Populus
11	Compositae	31	Quercus
12	Corylus	32	Rosaceae
13	Cupressus	33	Rumex
14	Cyperaceae	34	Salix
15	Chenopodium	35	Sambucus
16	Ericaceae	36	Ulmus
17	Fraxinus	37	Urticaceae
18	Helianthus	38	Urtica membranaceae
19	Juncaceae	40	No estudiado

Tabla A.1: Listado de tipos polínicos

DELETE. Una vez insertados todos los puntos del tipo insertado pulsar el botón de ENTER. Aparecerán cruces azules sobre los pólenes si la casilla “show” esta activa. Para añadir mas puntos, inserte el nuevo tipo, vuelva a pulsar ‘a’ y repita el proceso.

3) Eliminar un polen (DELETE): Si se ha añadido una cruz azul sobre algo que no es polen y se desea eliminar, pulse la tecla ‘d’ para cambiar la acción en la caja de texto correspondiente, sitúe el ratón encima de el punto y haga click. Se eliminará el punto.

b) Modo verificación: Permite borrar, cambiar tipo y añadir comentarios a la lista de polen anotado anteriormente (Ver apartado A.4) analizando imágenes a 40x de los tipos no identificados a 10x.

8. Generar fichero para analizar polen dudoso. El polen que no ha sido identificado, el de tipo 0, se analiza individualmente utilizando un enfoque de 40x aumentos. Para ello, se genera un fichero con el ID, la posición, enfoque y el diámetro del polen no identificado. Este fichero se guarda en el directorio donde se encuentra la imagen y se llama

\*NOMBRE-MUESTRA\*\_AUTOR\*\_XY\_Polen\_List\_No\_ID.dat.

### A.3. Digitalización de polen no identificado

Con la macro “FileXY\_polenList.mac” en NIS Elements se escanean los pólenes dudosos para un posterior análisis. Siga los siguientes pasos:

1. Coloque la muestra que desea estudiar.
2. Cambie el zoom a 40 aumentos.
3. Cambie en el software NIS a la configuración de 40 aumentos
4. Posicionar el objetivo con el software NIS Elements en el centro de la muestra, alrededor de  $X = 50000 \mu\text{m}$  e  $Y = 35200 \mu\text{m}$ .
5. Si la imagen se ve oscura, ajustar valores de “Exposure” .
6. Verifique que ha creado el fichero \*Autor\*\_XY\_Polen\_List\_No\_ID.dat que quiere analizar.
7. Ejecute la macro en Macro > Run from ... y seleccione “FileXY\_polenList.mac”.
  - a) Aparece una ventana con un contador a 0. Hay que utilizarlo cuando la captura no se ha finalizado completamente por fallo del hardware o software. Cuando esto suceda ir al directorio de la muestra analizada, al directorio \*Autor\*\_Z\_series y obtener el último punto semilla capturado a 40x correctamente (los puntos comienzan por este contador). En la macro añadir el número a partir del cual que quiere comenzar a realizar las capturas.
  - b) Ajuste de la posición X e Y. Debido a posibles variaciones al colocar la muestra en la platina y al movimiento que se produce al cambiar el tipo de aumento se permite modificar al inicio la posición de la imagen para ajustar el polen en el centro de la imagen. Pulse “YES” para modificar las posiciones y “NO” cuando estas sean correctas y continuar.
  - c) Ajuste del enfoque / posición Z. Debido a la posición relativa del eje Z hay que enfocar la muestra en el primer punto semilla. La posición de Z del resto de puntos semilla se calcula con la información del plano generado con los 49 puntos de enfoque a 10x y el desplazamiento llevado a cabo en este paso.
8. Una vez se han capturado todas las imágenes con puntos semilla con tipo desconocido (tipo 0), se puede verificar la muestra anotada por el autor para poner el tipo correspondiente.

## A.4. Verificación de puntos semilla anotados

En el análisis de la muestra a 10x se han añadido una serie de puntos, en algunos casos, sin tipo (tipo 0) que tienen que ser analizados posteriormente con imágenes a 40x para mejorar la anotación en una segunda fase. El modo de verificación permite borrar, cambiar tipo y añadir comentarios a la lista de polen anotado anteriormente. Para ello:

1. Seleccionar en el desplegable el autor a verificar.
2. Activar la casilla “CHECK” .
3. Añadir el nombre del autor que verifica y algún código identificador de lo que verifica( Ej: Si la anotación a revisar es Maria\_12-05-2012, el nombre de la revisión hecha por Pedro podría ser: Pedro\_Maria\_12-05-2012)
4. Seleccionar “SI” en la ventana que aparece, ya que queremos revisar la anotación hecha por el autor seleccionado.
5. Seleccionar “No ID” en la pregunta de si se desean revisar los pólenes no identificados o todos. En principio, se supone que si el tipo es distinto de cero es correcto por lo que solo verificamos los “No ID”.
6. A continuación aparece una ventana con la explicación de las teclas que se han de utilizar en la verificación. Hay tres funciones:

a) borrar

(pulsar tecla 'd')

b) cambiar tipo

(Nº en caja de tipo+ENTER+ {Comentario+ENTER(opcional)}+pulsar tecla 't')

NOTA1: El cambio de tipo también almacena el contenido de la caja de comentarios. Los comentarios añadidos junto al cambio de tipo pueden seguir la siguiente forma.

- ? 24 = Dudo entre el tipo puesto en la caja de tipo y el tipos 24
- ? = No estoy seguro del tipo (Solo poner cuando el tipo sea distinto de 0, de los no identificados)

NOTA2: Se ha definido el tipo 40 como polen no estudiado en este proyecto.

c) añadir comentarios

(Texto en caja de comentarios+ENTER+ pulsar tecla 'c') sin modificar el tipo. Para casos a 40x no identificados.

NOTA: Si a 40x no se identifica el tipo de polen añadir comentario de la forma:

- ? -1 23 = Dudo entre no polen (-1) o tipo 23
- ? 24 31 37 = Dudo entre tipos 24, 31 y 37

El objetivo con estos códigos es aprender de la imagen para un futuro sistema semiautomático mejorado.

7. Se recorren todos los pólenes de la anotación hasta que se finalice la lista o hasta que se pulse el botón ESC (salir de modo de verificación). Todas las acciones se guardan en la tabla de cambios. Si "Show 40x pollen" esta activo (por defecto), se abre la pila de imágenes capturadas a 40 aumentos en el punto semilla que se esta verificando en el software NIS Elements. Deslice la barra para analizar el tipo, descartarlo o añadir un comentario en la aplicación "PolenGUI". Cierre frecuentemente las diferentes imágenes que se van abriendo la pila para no saturar la memoria del sistema.

OJO: En la verificación no se puede ir hacia atrás, es muy importante que se realice la acción correcta ya que si no hay que eliminar el autor de verificación ("Delete") y comenzar de nuevo.





## Apéndice B

# Características analizadas de los granos de polen

A partir de la segmentación de objetos con características similares a los granos de polen, ya sea a partir de puntos semilla localizados manualmente por los expertos o a partir de los puntos localizados por el algoritmo automático de detección desarrollado, se analizan 67 características de color, tamaño, forma y textura.

Para el cálculo de las características se utiliza la imagen original y a máscara obtenida de la segmentación del objeto a partir de un punto semilla (punto verde en imagen derecha de figura B.1). También se analiza la distribución de la imagen en tres anillos concéntricos en función del eje mayor y del eje menor de la segmentación del grano de polen (ver imagen izquierda de figura B.1).

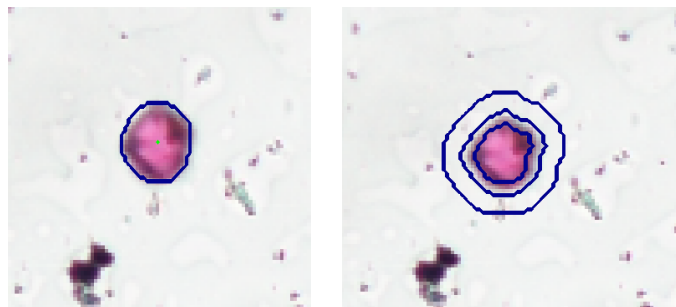


Figura B.1: Segmentación a partir de punto semilla y disposición de los anillos concéntricos para análisis de color y textura

A continuación se detalla cada una de las características visuales calculadas.

- Características de tamaño:
  - **Area:** Número de píxeles que componen la segmentación del objeto detectado.

- **EjeMayor**: Longitud del eje mayor de la segmentación de objeto detectado.
  - **EjeMenor**: Longitud del eje menor de la segmentación del objeto detectado.
  - **{Rojo, Azul, Verde}Medio**: Valor medio del canal {rojo, verde, azul} de los pixeles del objeto segmentado.
  - **{Rojo, Azul, Verde}Desv**: Valor de la desviación típica del canal {rojo, verde, azul} de los pixeles del objeto segmentado.
  - **{H, S, V}Medio**: Valor medio del canal {H, S, V} de los pixeles del objeto segmentado.
  - **{H, S, V}Desv**: Valor de la desviación típica del canal {H, S, V} de los pixeles del objeto segmentado.
  - **{A1, A2, A3}{RGB, HSV}Medio**: Valor medio del anillo {interno, intermedio, externo} de los pixeles del objeto segmentado para los diferentes canales en espacio {RGB, HSV}
  - **{A1, A2, A3}{RGB, HSV}Desv**: Valor de la desviación típica del canal {H, S, V} de los pixeles del objeto segmentado.
- Características de forma:
    - **Eccentricity**: Excentricidad del objeto segmentado. Un círculo tiene excentricidad 0 y un segmento lineal excentricidad 1.
  - Características de textura:
    - **{A1, A2, A3}Contrast**: Valor de contraste calculado con los pixeles del anillo {interno, intermedio, externo} a partir de la matriz de co-ocurrencia de niveles de gris<sup>1</sup> con tamaño de ventana 8 x 8 y vecindad de radio 1 y ángulo de 0°. Mide la variación local en una imagen.

$$CONTRASTE = \sum_{i,j} |i - j|^2 p(i, j) \quad Rango = [0 (Ng - 1)^2]$$

con  $Ng = NumNivelesGrises$

---

<sup>1</sup>La GLCM es una matriz de frecuencias con la que un pixel con un nivel de gris (i) aparece en una relación de espacio específica con otro pixel de nivel de gris (j) en función de la vecindad elegida y el ángulo (ver figura 3.11).

- **{A1, A2, A3}Correlation:** Valor de correlación calculado con los pixeles del anillo {interno, intermedio, externo} a partir de la matriz de co-ocurrencia de niveles de gris con tamaño de ventana 8 x 8 y vecindad de radio 1 y ángulo de 0°. Mide la similitud entre pixeles vecinos.

$$CORRELACION = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad Rango = [-1 \ 1]$$

- **{A1, A2, A3}Energy:** Valor de energía calculado con los pixeles del anillo {interno, intermedio, externo} a partir de la matriz de co-ocurrencia de niveles de gris con tamaño de ventana 8 x 8 y vecindad de radio 1 y ángulo de 0°. Mide la variabilidad de intensidad de un conjunto de pixeles.

$$ENERGY = \sum_{i,j} p(i, j)^2 \quad Rango = [0 \ 1]$$

- **{A1, A2, A3}Homogeneity:** Valor de homogeneidad local calculado con los pixeles del anillo {interno, intermedio, externo} a partir de la matriz de co-ocurrencia de niveles de gris con tamaño de ventana 8 x 8 y vecindad de radio 1 y ángulo de 0°. Mide la uniformidad de los niveles de grises en la imagen, es lo opuesto al contraste.

$$HOMOGENEITY = \sum_{i,j} \frac{p(i, j)}{1 + |i - j|} \quad Rango = [0 \ 1]$$

- **{A1, A2, A3}Entropy:** Valor de entropía calculado con los pixeles del anillo {interno, intermedio, externo} a partir de la matriz de co-ocurrencia de niveles de gris con tamaño de ventana 8 x 8 y vecindad de radio 1 y ángulo de 0°. Informa sobre la distribución de niveles de gris de la imagen.

$$ENTROPIA = \sum_i p(i) * \log_2(p(i))$$



# Apéndice C

## Resultados de los clasificadores analizados

En este apéndice se adjuntan los resultados obtenidos por el software Weka.

### C.1. Clasificador con todas las características analizadas

Se entrena un clasificador basado en arboles de decisión con las 67 características analizadas de los granos de polen en una muestra balanceada. Se obtiene el siguiente árbol de decisión.

==== Run information ====

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: VC\_PollenGrain-weka.filters.unsupervised.attribute.NumericToNominal-

Rlast-weka.filters.supervised.instance.SpreadSubsample-M1.0-X4941.0-S1

Instances: 9882

Attributes: 68

Area

EjeMayor

EjeMenor

Eccentricity

RojoMedio

RojoDesv

VerdeMedio

VerdeDesv

AzulMedio

AzulDesv

HMedio

HDesv

SMedio

SDesv

VMedio

VDesv

A1RojoMedio

A1RojoDesv

A1VerdeMedio

A1VerdeDesv

A1AzulMedio

A1AzulDesv

A1HMedio

A1HDesv

A1SMedio

A1SDesv

A1VMedio

A1VDesv

A2RojoMedio

A2RojoDesv

A2VerdeMedio

A2VerdeDesv

A2AzulMedio

A2AzulDesv

A2HMedio

A2HDesv

A2SMedio

A2SDesv

A2VMedio

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      8929      90.3562 %
Incorrectly Classified Instances    953      9.6438 %
Kappa statistic                    0.8071
Mean absolute error                0.1071
Root mean squared error            0.303
Relative absolute error            21.4286 %
Root relative squared error        60.5938 %
Total Number of Instances          9882

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.898   0.09    0.908     0.898   0.903     0.89     0
                0.91    0.102   0.899     0.91    0.904     0.89     1
Weighted Avg.   0.904   0.096   0.904     0.904   0.904     0.89

=== Confusion Matrix ===
      a  b  <-- classified as
4435 506 |  a = 0
447 4494 |  b = 1

```

Figura C.1: Resultados entrenamiento muestra balanceada con todas las características

En la figura C.1 se muestran los resultados obtenidos del modelo entrenado, con una tasa de verdaderos positivos del 91 % y una tasa de falsos positivos del 10.2 %. En la figura C.2 se muestran los resultados de evaluar toda la muestra con el modelo entrenado. Se obtiene una tasa de verdaderos positivos del 98.2 % y una tasa de falsos positivos del 9.1 %.

## C.2. Clasificador con el mejor subconjunto de las características analizadas

Se realiza una selección de los mejores atributos mediante la búsqueda de mejores subconjuntos con un algoritmo genético de 1000 individuos y un tamaño máximo de iteraciones de 2000. Se obtiene que el subconjunto que mejor se comporta tiene 16 características que se listan a continuación:

Con este subconjunto se atributos se entrena un clasificador basado en arboles de decisión con una muestra balanceada, obteniéndose el siguiente árbol:

```

=== Re-evaluation on test set ===
User supplied test set
Relation: VC_PollenGrain-weka.filters.unsupervised.attribute.NumericToNominal-Rlast
Instances: unknown (yet). Reading incrementally
Attributes: 68

=== Summary ===
Correctly Classified Instances      124385          91.2047 %
Incorrectly Classified Instances    11995           8.7953 %
Kappa statistic                    0.4145
Mean absolute error                 0.1002
Root mean squared error             0.2901
Total Number of Instances          136380

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
Weighted Avg.   0.909    0.018    0.999     0.909    0.952     0.942     0
                0.982    0.091    0.29      0.982    0.447     0.942     1
Weighted Avg.   0.912    0.02     0.974     0.912    0.934     0.942

=== Confusion Matrix ===
      a      b  <-- classified as
119531  11908 |      a = 0
   87    4854 |      b = 1

```

Figura C.2: Resultados evaluación a toda la muestra con el modelo entrenado con muestra balanceada y todas las características

```

==== Run information ====
Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: VC_PollenGrain-weka.filters.unsupervised.attribute.NumericToNominal-
Rlast-weka.filters.supervised.instance.SpreadSubsample-M1.0-
X4941.0-S1-weka.filters.unsupervised.attribute.Remove-V-
R1,4,5,11,17,23,33,35,47,51,54,56,57,61,63,64,68
Instances: 9882
Attributes: 17
Area
Eccentricity
RojoMedio
HMedio
A1RojoMedio
A1HMedio
A2AzulMedio
A2HMedio
A3HMedio
A3VMedio
A1Correlation
A1Homogeneity
A1Entropy
A2Homogeneity
A3Contrast
A3Correlation
Class
Test mode:10-fold cross-validation
==== Classifier model (full training set) ====

```



Selected attributes: 1,4,5,11,17,23,33,35,47,51,54,56,57,61,63,64 : 16

1. Area
2. Eccentricity
3. RojoMedio
4. HMedio
5. A1RojoMedio
6. A1HMedio
7. A2AzulMedio
8. A2HMedio
9. A3HMedio
10. A3VMedio
11. A1Correlation
12. A1Homogeneity
13. A1Entropy
14. A2Homogeneity
15. A3Contrast
16. A3Correlation

Figura C.3: Listado de los mejores atributos

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      8960           90.6699 %
Incorrectly Classified Instances    922           9.3301 %
Kappa statistic                    0.8134
Mean absolute error                0.116
Root mean squared error            0.2866
Relative absolute error            23.1954 %
Root relative squared error        57.3105 %
Total Number of Instances          9882

=== Detailed Accuracy By Class ===

                TP Rate   FP Rate   Precision   Recall   F-Measure   ROC Area
                0.905    0.091    0.908      0.905    0.907       0.919
                0.909    0.095    0.905      0.909    0.907       0.919
Weighted Avg.   0.907    0.093    0.907      0.907    0.907       0.919

=== Confusion Matrix ===

   a    b  <-- classified as
4470  471 |    a = 0
  451 4490 |    b = 1

```

Figura C.4: Resultados entrenamiento muestra balanceada con un subconjunto de las características

En la figura C.4 se muestran los resultados obtenidos del modelo entrenado, con una tasa de verdaderos positivos del 90.9% y una tasa de falsos positivos del 9.5%. En la figura C.2 se muestran los resultados de evaluar toda la muestra con el modelo entrenado. Se obtiene una tasa de verdaderos positivos del 95.8% y una tasa de falsos positivos del 8.4%.

```

=== Re-evaluation on test set ===

User supplied test set
Relation:      VC_PollenGrain-weka.filters.unsupervised.attribute.Numer
Instances:     unknown (yet). Reading incrementally
Attributes:    17

=== Summary ===

Correctly Classified Instances      125153          91.7679 %
Incorrectly Classified Instances    11227           8.2321 %
Kappa statistic                     0.4258
Mean absolute error                 0.1093
Root mean squared error             0.2761
Total Number of Instances          136380

=== Detailed Accuracy By Class ===

                TP Rate   FP Rate   Precision   Recall   F-Measure   ROC
                0.916     0.042     0.998       0.916    0.955       0
                0.958     0.084     0.3         0.958    0.457       0
Weighted Avg.   0.918     0.044     0.973       0.918    0.937       0

=== Confusion Matrix ===

      a      b  <-- classified as
120420 11019 |      a = 0
   208   4733 |      b = 1

```

Figura C.5: Resultados evaluación a toda la muestra con el modelo entrenado con muestra balanceada y un subconjunto de las características



## Apéndice D

### Imagen y características de los tipos polínicos identificados

Tras el análisis de 12 muestras procedentes de un captador de Toledo se han identificado 24 tipos polínicos. A continuación se presenta una imagen de cada uno de ellos junto a sus características más destacadas.

Tabla D.1: Imagen y características de los granos de polen identificados

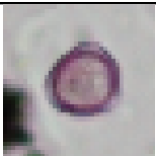

Imagen	Tipo	Nombre	Características	Num
	1	Acer	10-22 $\mu\text{m}$ isopolar Radiosimétrico Prolado Diporado Granuloso	3
	2	Alnus	13-28 $\mu\text{m}$ isopolar Radiosimétrico 5-zonocolporado aspidaperturado Protuberancias	2

Tabla D.1: Imagen y características de los granos de polen identificados





Imagen	Tipo	Nombre	Características	Num
	3	Apiaceae	14-25 $\mu\text{m}$ isopolar Radiosimétrico 3-zonocolporado Planaperturado Alas	6
	4	Artemisia	17-28 $\mu\text{m}$ isopolar Radiosimétrico trizonocolporado Ectoaberturas Equinulado-granulada	2
	5	Betula	20 - 25 $\mu\text{m}$ isopolar Esferoidal Monoica Triporado Granuloso	1
	10	Cedros	80-110 $\mu\text{m}$ heteropolar Radiosimétrico analeptomado Inaperturado sacci	1

Tabla D.1: Imagen y características de los granos de polen identificados


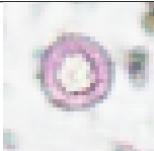
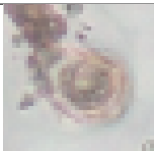
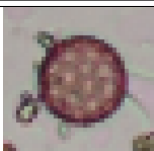
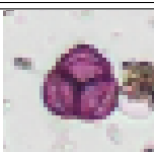
Imagen	Tipo	Nombre	Características	Num
	11	Compositae	16-20 $\mu\text{m}$ isopolar Radiosimétrico ectoapertura trizonocolporado Granuloso	11
	13	Cupressus	20-36 $\mu\text{m}$ esferoidal Radiosimétrico inaperturado Estrellada Lisa	723
	14	Cyperaceae	$\mu\text{m}$ isopolar Radiosimétrico Prolado Diporado Granuloso	2
	15	Chenopodium	10 -28 $\mu\text{m}$ esferoidal Radiosimétrico polipantoporado Ondulado Granuloso	23
	16	Ericaceae	20-40 $\mu\text{m}$ isopolar Radiosimétrico 4 celdas Triporado Reticulado	19

Tabla D.1: Imagen y características de los granos de polen identificados

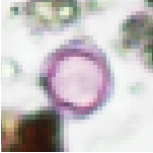

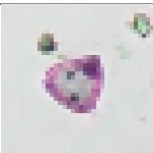


Imagen	Tipo	Nombre	Características	Num
	17	Fraxinus	18-23 $\mu\text{m}$ isopolar Radiosimétrico Suboblatos Triporado Reticulado	2
	22	Moraceae	7-17 $\mu\text{m}$ isopolar Radiosimétrico Prolado Diporado Granuloso	59
	23	Myrtaceae	20-25 $\mu\text{m}$ isopolar Radiosimétrico 3- zonosincolporado Endoaberturado Granuloso	2
	24	Olea	18-25 $\mu\text{m}$ Prolato-esferoidal Radiosimétrico Trizonicolporado Endoaberturado Granuloso	23
	26	Pinus	68-95 $\mu\text{m}$ Heteropolar Radiosimétrico analeptomado Inaperturado. Verrugoso	133



Tabla D.1: Imagen y características de los granos de polen identificados




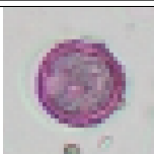

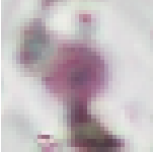


Imagen	Tipo	Nombre	Características	Num
	27	Plantago	22-44 $\mu\text{m}$ Apolar Radiosimétrico Verrugosa Pantoporado. 5-16 poros. Con opérculo y anillo	21
	28	Platanus	17-22 $\mu\text{m}$ Isopolar Radiosimétrico Trizonicolpado Planaperturado Granulosa	60
	29	Poaceae	20-100 $\mu\text{m}$ Heteropolar Simetría bilateral Granulada 1 poro Con opérculo y anillo	142
	30	Populus	29-35 $\mu\text{m}$ Isopolar Radiosimétrico Perforada Inaperturado Anillo de 2 $\mu\text{m}$	1012
	31	Quercus	22-30 $\mu\text{m}$ Isopolar Radiosimétrico oblato-esferoidal 1 poro Verrugoso	1030

Tabla D.1: Imagen y características de los granos de polen identificados

Imagen	Tipo	Nombre	Características	Num
	33	Rumex	17-20 $\mu\text{m}$ Isopolar tri otetrazonicolporados Penta- hexazoniporado 1 poro Granulado	3
	36	Ulmus	23-30 $\mu\text{m}$ Esferoidal Subcircular Penta- hexazoniporado Poros de 2 $\mu\text{m}$ Con báculos	25
	37	Urticaceae	12-17 $\mu\text{m}$ Isopolar Radiosimétrico Zopiporado 1 a 4 poros. Tectado	45





# Listado de referencias

- Allen, G. (2006). *An Automated Pollen Recognition System*. PhD thesis, Massey University.
- Barnes, C., Pacheco, F., Landuyt, J., Hu, F., and Portnoy, J. (2001). Hourly variation of airborne ragweed pollen in kansas city. *Annals of Allergy, Asthma & Immunology*, 86(2):166 – 171.
- Blackely, C. (1873). Experimental researches on the nature and causes of catarrhus aestivus. *Tindal & Cox*.
- Boucher, A., Hidalgo, P. J., Thonnat, M., Belmonte, J., Galan, C., Bonton, P., and Tomczak, R. (2002a). Development of a semi-automatic system for pollen recognition. *Aerobiologia*, 18:195–201.
- Boucher, A., Tomczak, R., Hidalgo, P., Thonnat, M., Belmonte, J., Galan, C., and Bonton, P. (2002b). Method and apparatus for the automatic detection and recognition of pollen.
- Branham, R. and Branham, D. (2012). Apparatus, system and method for automatic airborne contaminant analysys.
- Durham, O. C. (1946). The volumetric incidence of airborne allergens. *J. Allergy*, 17:79–86.
- Flenley, J. (1968). The problem of pollen recognition. *Problems in Picture Interpretation*, pages 141–145.
- France, I. and Duller, A. (1999). Beyond paradise: Extensions to a neural network for visual object recognition. In Singh, S., editor, *International Conference on Advances in Pattern Recognition*, pages 206–215. Springer London.
- France, I., Duller, A., Duller, G., and Lamb, H. (2000). A new approach to automated pollen analysis. *Quaternary Science Reviews*, 19:537–546.

- Galán, C., García, H., Cariñanos, P., Alcazar, P., and Domínguez, E. (2007). *Spanish Aerobiology Network (REA) Management and Quality Manual*. Servicio de publicaciones de la Universidad de Córdoba.
- Haralick, R., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(6):610–621.
- Hirst, J. (1952). An automatic volumetric spore trap. *Ann. Appl. Biol.*, 39:257–265.
- Hodgson, R. M., Holdaway, C., Zhang, Y., Fountain, D., and Flenley, J. (2005). Progress towards a system for the automatic recognition of pollen using light microscope images. In *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on*, pages 76–81.
- Holt, K., Allen, G., Hodgson, R., Marsland, S., and Flenley, J. (2011). Progress towards an automated trainable pollen location and classification system for use in the palynology laboratory. *Review of Palaeobotany and Palynology*, 167:175–183.
- Holt, K. A. and Bennett, K. D. (2014). Principles and methods for automated palynology. *New Phytologist*, 203:735–742.
- Jones, G. D. (2012). Pollen analyses for pollination research, unacetoyzed pollen. *Journal of Pollination Ecology*, 9(12):96–107.
- Landsmeer, S. H., Hendriks, E. A., Weger, L. A., Reiber, J. H. C., and Stoel, B. C. (2009). Detection of pollen grains in multifocal optical microscopy images of air samples. *Microscopy Research and Technique*, 72:424–430.
- Prieto, R. A., Santosjuanes, I. N., Gómez, M. R., and Ruiz, J. F. S. (2011). Reconocimiento de granos de polen. Technical report, Visión Artificial. UNED.
- Puc, M. (2012). Influence of meteorological parameters and air pollution on hourly fluctuation of birch (*Betula L.*) and ash (*Fraxinus L.*) airborne pollen. *ANNALS OF AGRICULTURAL AND ENVIRONMENTAL MEDICINE*, 19(4):660–665.
- Ranzato, M., Taylor, P., House, J., Flagan, R., LeCun, Y., and Perona, P. (2007). Automatic recognition of biological particles in microscopic images. *Pattern Recognition Letters*, 28(1):31 – 39.
- Rincón, M., Hope, T., Bjornerud, A., Selnes, P., and Fladby, T. (2011). Automatic wml segmentation and quantification using a machine learning approach. *The 19th International Society for Magnetic Resonance in Medicine Annual Meeting*.

- Rodriguez-Damian, M., Cernadas, E., Formella, A., Fernandez-Delgado, M., and Sa-Otero, P. D. (2006). Automatic detection and classification of grains of pollen based on shape and texture. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 36(4):531–542.
- Ronneberger, O. (2007). *3D Invariants for Automated Pollen Recognition*. PhD thesis.
- Sikoparija, B., Pejak-Sikoparija, T., Radisic, P., Smith, M., and Galan, C. (2011). The effect of changes to the method of estimating the pollen count from aerobiological samples. *Journal of Environment Monitoring*, 13(2):384–90.
- Stillman, E. and Flenley, F. J. (1996). The needs and prospects for automation in palynology. *Quaternary Science Reviews*, 15:1–5.
- Subiza, J. (2001). Cómo interpretar los recuentos de pólenes. *Alergol Inmunol Clin*, 16:59–65.

